Psy.

DECEMBER
1 9 5 1
VOLUME 15
NUMBER 6

UNIVERSITY
OF MICHIGAN
FEB 18 1952
SCIENCE
LIBRARY

# JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

# JOURNAL OF CONSULTING PSYCHOLOGY

EDITED BY LAURANCE F. SHAFFER

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

ASSOCIATE EDITORS: EDGAR A. DOLL, *Devereux Schools, Devon, Pa.* WILLIAM A. HUNT, *Northwestern University.* E. LOWELL KELLY, *University of Michigan,* MORRIS KRUGMAN, *Board of Education, New York, N. Y.* BERTHA M. LUCKEY, *Board of Education, Cleveland, Ohio.* FRED MCKINNEY, *University of Missouri.* ANNE ROE, *New York, N. Y.* CARL R. ROGERS, *University of Chicago.* R. NEVITT SANFORD, *University of California.*

MARGARET K. HARLOW, *Managing Editor*

THE JOURNAL OF CONSULTING PSYCHOLOGY is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

## EDITORIAL OFFICE

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

*Information for contributors*: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

Some of the mechanical requirements for manuscripts are: Type manuscripts with wide margins, on white bond paper. Double space between *each and every line,* including references, footnotes, matter to be set in smaller type, and tables. Use only two types of headings: centered headings, and underlined paragraph headings. Type each table, however small, on a separate sheet, with Arabic number and title. Draw all figures in India ink in final form for reproduction, with draftsman's lettering, but type captions for figures on separate sheets. The *References* must follow the exact form used in APA journals. For further information on the details of preparing manuscripts, see:

> ANDERSON, J. E., AND VALENTINE, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.,* 1944, **14,** 345-376.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $14.00 per printed page.

Authors may order reprints when returning proof. No reprints are furnished gratis for articles submitted after January 1, 1952. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

## BUSINESS OFFICE

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters should be sent to the American Psychological Association, Inc., 1515 Massachusetts Ave., N. W., Washington 5, D. C.

The Journal of Consulting Psychology is published bi-monthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $5.00, foreign $5.50, single number $1.00.

Copyright 1951 by the American Psychological Association, Inc. Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

December, 1951 Vol. 15, No. 6

## CONTENTS

A Factorial Analysis of Fifteen MMPI Scales: FRED T. TYLER - - - - - - - - - 451

The Humm-Wadsworth and the Minnesota Multiphasic: A. R. GILLILAND - - - - 457

The Rorschach and Descriptive Diagnosis: J. R. WITTENBORN AND JULES D. HOLZBERG - - - - - - - - - - - - - - - - - - - - - 460

Impulsivity as Indicated by Rorschach Test Factors: RILEY W. GARDNER - - - - 464

Experimentally Induced Faking of TAT Responses: EDITH A. WEISSKOPF AND JORGE J. DIEPPA - - - - - - - - - - - - - - - - - - - - - - - - 469

Pupil Functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children: JUDITH I. KRUGMAN, JOSEPH JUSTMAN, J. WAYNE WRIGHTSTONE, AND MORRIS KRUGMAN - - - - - - - - - - - - - - - - - - - - - - - - 475

A Comparison of Masculinity-Femininity Measures: BERNARD F. SHEPLER - - - - 484

The Semantics of Intellectual Classification: RAYMOND CORSINI - - - - - - - - 487

Psychological Tests of Research Scientists: ANNE ROE - - - - - - - - - - - - 492

Self-Evaluating Questionnaires as Projective Measures of Personality: GABRIEL ELIAS - 496

A Psychometric Study of Drinkers: PETER J. HAMPTON - - - - - - - - - - - - 501

Practical Correlates of Psychiatric Symptoms: J. R. WITTENBORN AND FRED A. METTLER 505

A Note on Fosberg's "Four Experiments with the Szondi Test": JACOB COHEN - - - 511

New Books and Tests - - - - - - - - - - - - - - - - - - - - - - - - - - 512

Index to Volume 15 - - - - - - - - - - - - - - - - - - - - - - - - - - - 517

# A FACTORIAL ANALYSIS OF FIFTEEN MMPI SCALES

FRED T. TYLER[1]
UNIVERSITY OF CALIFORNIA

IT is only quite recently that any reports of factor analyses of the Minnesota Multiphasic Personality Inventory have been published. Marjory Stout [12] reported a factor analysis of the original nine scales of the MMPI for college men and for college women. The results showed that four factors accounted for the major part of the interrelationships for each sex. Actually, the loadings on all but the first factor were rather small, only two in the second and one in each of the third and fourth factors exceeding .30. Cook and Wherry [1], using 120 naval enlisted submarine personnel as subjects, analyzed a battery of 20 variables, including 11 from the MMPI. Six factors were isolated, loadings of .20 or larger being regarded as significant. The first factor included all the MMPI scales with the exception of the *L* and *F* scales. This factor was labeled the "tendency to personality maladjustment." The third factor had positive loadings on the *Ma* scale and negative loadings on *L*, *Pd*, *Hs*, *D*, and *Hy*. The latter three are frequently referred to as the "neurotic triad." The whole factor was titled "tendency to overactivity." The factor was interpreted as involving lack of self-consciousness and self-criticism, with a tendency toward impatience and impulsiveness. Factor IV included *L*, *Hy*, *Mf*, and *Pa*, and was labeled "tendency to paranoia," because the largest loading, .48, was with *Pa*. The sixth factor was entitled "tendency to femininity of interest pattern." The loadings were positive for *Mf* and *Pt*, and negative for *Hy*. The suggestion here was that the more effeminate men tend toward compulsive behavior.

Cottle [2] reported a factor analysis of the MMPI, the Strong Vocational Interest Blank for Men, the Kuder Preference Record, and the Bell Adjustment Inventory. The subjects were 400 male veterans who had been referred to the Psychological Services Center of Syracuse University. The factors were extracted by Thurstone's complete centroid method, the final primary factors being based upon an oblique rotation of the axes. The MMPI contribution to each factor is summarized:

1. Factor B included *F*, *Hs*, *Pd*, *Pt*, *Sc*, and *Ma*, with *L* having a negative loading. The largest loadings were contributed by the scales included in the "psychotic phase" of the MMPI. The factor was labeled "more serious maladjusted tendencies."
2. Factor D consisted of *Hs*, *D*, and *Hy*, the scales included in the neurotic triad. The factor was considered to indicate "neurotic tendencies."
3. Factor E included the *Mf* scale (and the Kuder MF, and was considered to be a "residual plane."

In general, Cottle found practically no overlapping between personality and interest inventories.

Wheeler, Little, and Lehner [14] investigated the internal structure of the MMPI using two groups of subjects, male college students and male neuropsychiatric patients in a veterans hospital. Their correlation matrix included twelve variables, the nine clinical scales and the *L*, *F*, and *K* scales. Two major and two minor factors emerged from Thurstone's centroid multiple factor analysis; the interpretation is based upon the factors obtained after rotation. Reference will here be confined to the results obtained from college men. Factor I had its greatest loadings on the *Sc* and *Pt* scales, components of what is usually known as the "psychotic triad." The writers believe this factor indicates concern with one's self and reflects "withdrawal of a schizoid type including excessive concern with compulsive needs." The second factor had its maximum loading on *Hy* and high loadings on *Hs* and

[1]The writer is indebted to Harrison G. Gough for scoring stencils for several scales here used and for assistance in interpreting the factors.

*D*, the three components of the neurotic triad. The second factor is believed to reveal a picture of neurotic behavior. The third and fourth factors are difficult to interpret inasmuch as each includes only two scales. One of these includes the *Mf* and the *Pa* scales, reflecting a possible relationship between masculinity-femininity and paranoia in normal male subjects.

## PROBLEM

The purpose of the present paper is to report the results of a factor analysis of the MMPI scores of 107 female graduate students preparing to become elementary teachers at the University of California, Berkeley. All subjects had passed a physical examination, and none had been referred to a psychiatrist for further examination. The subjects answered all questions in the booklet form of the MMPI. This made it possible to obtain scores for 15 scales exclusive of the *F*, *L*, and Validity scales. In addition to the original 9 scales, scores were obtained for:

1. *Si*: a scale for social introversion [3, 5].
2. *St*: a scale for social status [8].
3. *Pr*: a measure of prejudice [7].
4. *Ac*: a measure of academic achievement drive [6].
5. *Re*: a scale prepared to measure social responsibility [7].
6. *Do*: a measure of social dominance [7].

Table 1 shows the mean scores (raw and *T*) for the group of women used in the present analysis. It may be noted from Table 1 that there is a distinct tendency, with a few exceptions, for the group means to center closely on a *T*-score of 50. The group is somewhat below average on Hypochrondriasis and on Prejudice. On the other hand, the mean scores are larger than 50 on Status, Responsibility, and Dominance. Further data may reveal whether these divergencies are characteristic of college students in general or more particularly of women students planning to become elementary school teachers. Certainly this group does not reveal any marked tendency toward psychoneuroses or neuroses as defined by the elevated scores on the triads of *Pt*, *Sc*, and *Pa*, and *Hs*, *D*, and *Hy*, respectively.

TABLE 1

MEANS AND SIGMAS OF SUBJECTS ON EACH SCALE: MEAN *T*-SCORES FOR THREE STUDIES

| Subtest | Description of Subtest | Raw | | T-score* | Stout's Subjects |
|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *M* |
| *Hs* | Hypochondriasis | 3.56 | 2.64 | 44 | 46 |
| *D* | Depression | 19.46 | 4.29 | 50 | 51 |
| *Hy* | Hysteria | 20.99 | 3.46 | 54 | 55 |
| *Pd* | Psychopathic Deviate | 13.02 | 3.58 | 49 | 53 |
| *Mf* | Masculinity-Femininity | 38.28 | 3.66 | 47 | 49 |
| *Pa* | Paranoia | 9.52 | 2.64 | 54 | 53 |
| *Pt* | Psychasthenia | 9.88 | 5.43 | 46 | 49 |
| *Sc* | Schizophrenia | 9.27 | 6.09 | 48 | 50 |
| *Ma* | Hypomania | 14.94 | 3.36 | 52 | 57 |
| *Si* | Social Introversion | 25.04 | 7.55 | 50 | |
| *St* | Status | 24.02 | 3.11 | 64 | |
| *Pr* | Prejudice | 6.39 | 3.99 | 37 | |
| *Ac* | Academic Achievement | 17.37 | 2.63 | 51 | |
| *Re* | Responsibility | 24.41 | 2.32 | 63 | |
| *Do* | Dominance | 19.64 | 2.44 | 63 | |

*T-scores for the original nine scales were obtained from the Manual of Directions; T-scores for the remaining scales were obtained from data supplied by Gough.

TABLE 2

DESCRIPTIONS OF MMPI SCALES

| | |
|---|---|
| *Hs* | Concern over bodily functions, concern about health, tendency toward physical complaint |
| *D* | Depression, dejection, discouragement, despondency |
| *Hy* | Immature, unrealistic, amenable to group ideas, kindly, courteous, naive, needs social acceptance |
| *Pd* | Irresponsible, undependable, impulsive, egocentric, defiant, asocial, individualistic |
| *Pa* | Aggressive, critical, irritable, moody, sensitive, sensitive to criticism |
| *Pt* | Apprehensive, tense, hesitant, insecure, self-conscious, feelings of inadequacy |
| *Sc* | Bashful, withdrawn, oversensitive, secretive, cautious |
| *Ma* | Confident, hypersensitive, not persistent, aggressive, charming, expansive |
| *Mf* | Masculinity or femininity of interests |
| *Pr* | Cynicism, poor morale, suspiciousness, feelings of ill will, lack of faith in others |
| *Si* | Tendency to avoid social contacts, little dependency upon people |
| *Ac* | Tenseness, docility, insecurity (producing greater academic output) |
| *St* | Self-confidence, poise, resourcefulness |
| *Re* | Dependability, integrity, seriousness, steadiness, tenacity |
| *Do* | Optimism, persuasiveness, self-discipline, resoluteness |

The present group of women had lower mean scores on all scales than did the subjects used by Stout except on *Mf* and *Pa*. Especially interesting are the differences on the *Hs*, *Pt*, *Sc*, and *Ma* scales, suggesting that in Stout's group there was more concern over physical well-being, more tenseness, insecurity and caution, and more social aggressiveness. If differences in MMPI scores continue to be reported from different universities, research workers may need to consider more carefully the problems arising from the nature of the samples used in an investigation.

Brief descriptions and synonyms for the various scales are shown in Table 2. These descriptions were obtained from the Manual of Directions for the MMPI, from Rosen's Ph.D. dissertation [11], from Drake [3], from Evans and McConnell [5], and from Gough [6, 7, 8].

Table 3 shows the intercorrelations, computed by machine from raw scores, for the 15 variables used in the present study. The authors of the MMPI indicate that the *L* score may have some value as a measure of personality trend, but they provide no interpretation. They say specifically that the "*F* score is not a personality scale." The manual for the *K* scale states: "*K* is not known to have much clinical significance in itself." The correction factor was not applied to the relevant scales because there is little evidence to indicate either greater reliability or validity for the corrected scores. Statistical studies of the *K* scale are in process, and it is hoped that the results may be published at an early date. It appears that an analysis based upon *K* scores and scores on corrected clinical scales would offer difficulty in interpretation inasmuch as five of the clinical scales would be correlated with *K* because *K* is part of the clinical score. For these reasons *L*, *F*, and *K* were not used in the present analysis. The correlations ranged from −.56 (*Ma* and *Re*) through zero up to .77 (*Pt* and *Sc*). The upper half of the table contains the intercorrelations. The correlation matrix was analyzed by Thurstone's complete centroid method. The analysis was carried through until five factors had been extracted, at which time the residuals varied from −.12 to + .11, with a mean of .04 when true communalities had been used in the analysis. The residuals are shown in the lower half of Table 3.

The resulting factorial patterns were rotated so as to produce both orthogonal and oblique axes. Unrotated and rotated factors (oblique and orthogonal) are shown in Tables 4, 5, 6, 7, and 8, together with a brief interpretation of each factor. The descriptions tend to be eclectic, being based upon rotated and unrotated factors. The three sets of loadings are presented so that the reader who prefers one of these factorial patterns to another may make his own interpretation.

Factor I (Table 4) includes 13 of the 15 variables and so is probably best considered as a "general maladjustment factor." In

TABLE 3

INTERCORRELATIONS AND RESIDUALS AFTER EXTRACTION OF FIVE FACTORS

| | *Hs* | *D* | *Hy* | *Pd* | *Mf* | *Pa* | *Pt* | *Sc* | *Ma* | *Si* | *St* | *Pr* | *Ac* | *Re* | *Do* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Hs* | | 49 | 24 | 34 | 05 | 08 | 55 | 35 | 15 | 39 | −18 | 18 | 45 | −07 | −13 |
| *D* | 02 | | 05 | 37 | 25 | 04 | 57 | 36 | −08 | 60 | −19 | 30 | 37 | −02 | −18 |
| *Hy* | 07 | −03 | | 16 | 10 | 29 | −07 | 02 | −16 | −23 | 17 | −32 | 02 | 28 | 13 |
| *Pd* | −09 | 07 | −01 | | 01 | 27 | 43 | 43 | 48 | 10 | −51 | 15 | 18 | −11 | 14 |
| *Mf* | 00 | −02 | 04 | 05 | | 07 | 17 | 21 | −35 | 11 | 28 | 12 | 23 | −04 | 11 |
| *Pa* | −01 | 08 | 03 | 07 | −01 | | 09 | 11 | 04 | −28 | 18 | −38 | −16 | 03 | 14 |
| *Pt* | 01 | −02 | −06 | 05 | 00 | 08 | | 77 | 27 | 59 | −40 | 48 | 49 | −19 | −36 |
| *Sc* | −09 | −08 | −02 | 03 | 04 | 03 | 06 | | 31 | 34 | −21 | 42 | 34 | −09 | −06 |
| *Ma* | −02 | −05 | −05 | 01 | −12 | −03 | 06 | 04 | | −11 | 05 | 28 | 18 | −56 | 02 |
| *Si* | 01 | 05 | −10 | 04 | −01 | 00 | 04 | 01 | −02 | | −51 | 52 | 41 | −01 | −36 |
| *St* | 05 | 00 | 01 | −07 | 01 | −03 | 00 | −03 | 05 | −02 | | −44 | −28 | 07 | 50 |
| *Pr* | 02 | −02 | 03 | 03 | 10 | −03 | 04 | 10 | 01 | −02 | −06 | | 33 | −34 | −27 |
| *Ac* | 07 | −08 | 08 | −04 | 10 | −05 | −03 | −05 | 07 | −09 | 01 | −07 | | −09 | −31 |
| *Re* | −10 | −06 | −02 | 04 | −06 | −06 | 02 | 11 | −05 | 01 | 04 | 06 | 03 | | 10 |
| *Do* | −03 | 03 | −05 | 05 | −01 | −04 | −03 | 09 | −06 | 05 | 04 | 07 | 08 | 06 | |

TABLE 4

FACTOR 1: UNROTATED AND ROTATED

| | Unrotated | Oblique | Orthogonal |
|---|---|---|---|
| *Hs* | 500 | 386 | 386 |
| *D* | 585 | 592 | 592 |
| *Pd* | 376 | | |
| *Mf* | | | |
| *Pa* | | –340 | –338 |
| *Pt* | 878 | 697 | 698 |
| *Sc* | 615 | 421 | 423 |
| *Ma* | 352 | | |
| *Si* | 719 | 849 | 847 |
| *St* | –489 | –538 | –538 |
| *Pr* | 739 | 702 | 702 |
| *Ac* | 595 | 562 | 561 |
| *Re* | –363 | | |
| *Do* | –379 | –460 | –459 |

normal subjects it should probably not be interpreted as measuring either "general neuroses" or "general psychoneuroses" even though loadings on *Hs, D, Pt,* and *Sc* are rather high. It appears, rather, to reflect a chronic state of marginal maladjustment including concern with one's self. Subjects who score high on the factor would tend to have difficulties in their social contacts, possibly by reason of their dissatisfactions, worry and disgruntlement, and their complaining and pessimistic attitude. Low scores would be indicative of a more socially poised, self-confident, and optimistic attitude. The factor is similar to Wheeler's Factor I, as far as the heavy loadings are concerned.

TABLE 5

FACTOR II: UNROTATED AND ROTATED

| | Unrotated | Oblique | Orthogonal |
|---|---|---|---|
| *Hs* | 399 | | (292) |
| *D* | 443 | 309 | 352 |
| *Hy* | 488 | 340 | 387 |
| *Pd* | 325 | | |
| *Mf* | 456 | | |
| *Pa* | 399 | 427 | 411 |
| *Pt* | (283) | 453 | 437 |
| *Sc* | (289) | | 304 |
| *Ma* | (–298) | –306 | –318 |
| *Pr* | –312 | (–299) | –316 |
| *Re* | 414 | 482 | 489 |

Factor II (Table 5) might be termed "a conflict factor" because it seems to suggest a conflict between opposing tendencies, together with a lack of insight. Dissatisfaction appears to be present in the various scales contributing to the factor, but probably to a lesser degree than in the preceding factor; with this dissatisfaction is a tendency toward excessive self-control. High scores would suggest the presence of a series of contrasts or conflicting attitudes: (1) need for independence, but with underlying dependency; (2) a facade of altruism and selflessness, along with a basic narcissism; (3) an emphasis on objectivity and frankness along with a heightened sensitiveness and rigidity; (4) self-pity along with fantasies of fortitude and resoluteness. The factor points to a type of neurotic (*Hs, D,* and *Hy*) personality with repressive, hysteroid characteristics. The general picture presented by the unrotated factor is similar to Wheeler's second factor.

TABLE 6

FACTOR III: UNROTATED AND ROTATED

| | Unrotated | Oblique | Orthogonal |
|---|---|---|---|
| *Hs* | | 317 | 331 |
| *D* | | | |
| *Hy* | | 385 | |
| *Pd* | 480 | 578 | 602 |
| *Mf* | | | |
| *Pa* | 398 | 557 | (266) |
| *Pt* | | | 548 |
| *Sc* | | 356 | 501 |
| *Ma* | 771 | 442 | 843 |
| *Si* | –451 | –331 | |
| *St* | | 328 | |
| *Pr* | | (–282) | |
| *Ac* | | | |
| *Re* | –349 | | –482 |
| *Do* | | 339 | |

The general picture of Factor III (Table 6) suggests a tendency toward social aggressiveness and consequent friction. Women scoring high on this factor would probably be somewhat inconsiderate of others, abrupt, irritable, and impatient in social contacts, and determined, impulsive, and resentful in their behavior. They are not likely to have any very satisfactory insight into their own role in producing their difficulties. They may reveal rather egocentric and hostile attitudes, but there is little likelihood of any real self-dissatisfaction. Such behavior as that outlined may make for unhappiness in social situations,

but a person with such characteristics could attain satisfactory social and professional goals if other traits and abilities were favorable.

TABLE 7

FACTOR IV: UNROTATED AND ROTATED

| | Unrotated | Oblique | Orthogonal |
|---|---|---|---|
| *Mf* | 347 | 545 | 544 |
| *Pt* | −359 | | |
| *St* | 418 | 483 | 482 |
| *Do* | | 328 | 328 |

In general, Factor IV (Table 7) indicates good social adjustment accompanied by strong drives and self-confidence. A woman scoring high on this factor is likely to be objective and open-minded in her behavior toward others. Women scoring high on this factor should also be characterized by forcefulness and efficiency in general.

TABLE 8

FACTOR V: UNROTATED AND ROTATED

| | Unrotated | Oblique | Orthogonal |
|---|---|---|---|
| *Hs* | 337 | 334 | 337 |
| *Mf* | −525 | −349 | −525 |

The nature of Factor V (Table 8) is not too well revealed because of the limited number of variables contributing to the factor. On the other hand, its interpretation is relatively clear-cut; the association of conventional "femininity" (weakness, helplessness, and dependency) with a distinct tendency toward physical complaints. Wheeler, Little, and Lehner found *Mf* to be related to *Pa* in college men; in the case of men, a tendency toward femininity is related to aggressive, critical, sensitive behavior.

## SUMMARY AND CONCLUSIONS

1. The MMPI was administered to 107 women students training at the University of California to become elementary school teachers. Fifteen scores were obtained for each student, the nine original clinical scales of the MMPI and six which have been developed more recently.

2. The intercorrelation matrix was factor-analyzed by the complete centroid method; the unrotated factors were compared with the factors obtained from both oblique and orthogonal axes.

3. Five factors were obtained, and descriptions were attempted. In general, the results are in agreement with interpretations of the MMPI by clinical psychologists.

4. Inasmuch as this is the only factor analysis available for the fifteen scales used in this study, it is not possible to compare very specifically the results with those of other investigators. Stout extracted four factors from the original nine scales of the MMPI, but in the main she found a heavy general factor running through these scales. This may be comparable to Factor I described in this study. Wheeler, Little, and Lehner reported two more or less well-defined group factors in the original nine clinical scales and the *L*, *F*, and *K* scales.

*Received February 21, 1951.*

## REFERENCES

1. COOK, E. B., AND WHERRY, R. J. A factor analysis of the MMPI and aptitude test data. *J. appl. Psychol.*, 1950, **34**, 260-266.
2. COTTLE, W. C. A factorial study of the Multiphasic, Strong, Kuder, and Bell inventories using a population of adult males. *Psychometrika*, 1950, **15**, 25-47.
3. DRAKE, L. E. A social IE scale for the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 51-54.
4. EVANS, CATHERINE, AND MCCONNELL, T. R. A new measure of introversion-extroversion. *J. Psychol*, 1941, **12**, 111-124.
5. EVANS, CATHERINE, AND MCCONNELL, T. R. *Minnesota T-S-E Inventory*. Chicago: Science Research Associates, 1942.
6. GOUGH, H. G. Factors relating to the academic achievement of high school students. *J. educ. Psychol.*, 1949, **40**, 65-78.
7. GOUGH, H. G. Mimeographed materials at the University of California, including preliminary *T*-scores for the following scales: *Re*, *Do*, *Pr*, *Ac*, and *St*.
8. GOUGH, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401-409.
9. GOUGH, H. G., MCCLOSKY, H., AND MEEHL, P. E. A personality scale for dominance. *J. abnorm. soc. Psychol.*, 1951, **46**, 360-366.
10. GOUGH, H. G., MCCLOSKY, H., AND MEEHL, P. E. A scale for measuring social responsibility. *J. abnorm. soc. Psychol.*, in press.

11. ROSEN, E. Some personality and attitude differences between volunteers and non-volunteers for a psychological experiment. Unpublished Doctor's dissertation, Univ. of Calif., 1950.
12. STOUT, MARJORY. An analysis of the structure of the Minnesota Multiphasic Personality Inventory. Unpublished Master's thesis, Pennsylvania State College, 1949.
13. THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univ. of Chicago Press, 1947.
14. WHEELER, W. M., LITTLE, K. B., AND LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134-141.

# THE HUMM-WADSWORTH AND THE MINNESOTA MULTIPHASIC

A. R. GILLILAND
NORTHWESTERN UNIVERSITY

THE present study consists of a comparison of two well-known personality tests, the Humm-Wadsworth Temperament Scale and the Minnesota Mulitphasic Personality Inventory. These tests were given to a group of college students for an interpretation of their personality characteristics. Because of the nature of the use of these tests in this study, the results bear only indirectly on their use for industrial or clinical purposes.

The tests were generally given to classes of about fifty advanced students made up largely of those majoring in psychology, education, speech, or journalism. Both the individual and group forms of the MMPI were used on some groups. All the tests were scored as directed in the latest manuals. The raw scores on each test were converted into their respective types of standard scores.

The Humm-Wadsworth and the MMPI each measure five components or traits which have the same or very similar names or descriptions. These are: (1) hysteroid—psychopathic deviate, (2) mania—hypomania, (3) depression, (4) autistic—schizoid, and (5) paranoia.

It has been pointed out, however, by one of the authors[1] of one of these tests that although identical or similar terminology has been used in naming some of the traits in the two tests, this does not necessarily mean that the factors measured are identical or even very similar. For example, in the Humm-Wadsworth scale normality or abnormality is largely measured by the first component, or at least its control effect over the other traits is indicated. No such method is used in the MMPI. High scores and seemingly low scores (although little mention is made of these) indicate abnormality on the MMPI. Of the five traits, probably there is the greatest difference between hysteroid and psychopathic deviate. Psychopathic deviate is the broader term, whereas hysteroid refers primarily only to antisocial or criminal tendencies. Nevertheless, there is enough common ground covered by the two terms to include them in the comparison of traits. Likewise, autistic is not as inclusive a term as schizophrenia. But the autistic tendency is certainly one of the outstanding characteristics of schizophrenia, and it seems reasonable to assume a high relationship between them.

No attempt was made, as suggested in the manual, to use the "normal" scores as a basis for interpreting the other components on the Humm-Wadsworth scale. This was partly due to the fact that the component control measure did not seem especially suited for this purpose. There was also another reason. Seemingly, each component of the Humm-Wadsworth was originally validated against an independent criterion. If so, this would justify the interpretation of scores on each part of the test as a measure of this component in the subject's personality.

Table 1 gives the Pearsonian correlations be-

TABLE 1

CORRELATIONS BETWEEN HUMM-WADSWORTH TEST AND MMPI

| Number of cases | Hysteroid-Psych. Deviate | Mania-Hypomania | Depression | Autistic-Schizoid | Paranoia |
|---|---|---|---|---|---|
| 45* | .08 | −.15 | .16 | .00 | −.19 |
| 127* | .10 | .20 | .07 | −.02 | −.05 |
| 23* | .00 | .11 | .25 | .21 | .00 |
| 34* | .08 | −.03 | .16 | .15 | .13 |
| 50† | .12 | .03 | — | — | −.24 |
| 50† | .00 | −.21 | .30 | .10 | −.06 |

*Group was administered the MMPI Group Form.
†Group was administered the MMPI Individual Form.

[1]In a personal letter from Dr. Doncaster G. Humm, dated December 30, 1949.

tween the five common components of the two tests for six different classes containing a total of 329 students. The first four lines of the table present correlations between the Humm-Wadsworth and the group MMPI for a total of 229 students. The last two groups of 100 students give a comparison between the individual form of the MMPI and the Humm-Wadsworth tests. No complicated statistical devices were used in interpreting the meaning of these correlations. This omission was not due to a failure to consider such possibilities but, rather, because it appeared that the best method of interpretation would come from an inspection of the data to determine trends or a lack of trends.

A study of the table of correlations shows almost exactly a chance distribution from zero as the mean. No general agreement between scores on the two tests can be found. This fact seems to hold for comparisons of the Humm-Wadsworth with both group and individual forms of the MMPI. The general statement may need to be qualified somewhat by saying that it holds true for at least three of the common traits, but possibly there is a slight general agreement between depression scores on the two tests, as shown by low positive correlations, and a slight consistent disagreement on paranoia, as shown by four out of six low, negative correlations. These results may be restated by saying that there is a slight tendency for a person making a high score on one to make a low score on the other for paranoia; and there is no agreement on the other three components.

The second part of this study consists of a comparison of each test with ratings on the five personality traits. Three ratings on each of the five traits were obtained for each subject. One of these was a self-rating on brief descriptions of each trait. The outlines of the traits were constructed by a comparison of the traits as described in the two manuals. Two additional ratings were obtained from close associates of the subjects who were well trained in the technique of personality rating. The ratings of friends were returned in sealed envelopes without the results being made known to the persons being rated.

Table 2 gives the correlations of the self-ratings with the means of the ratings by two associates for each of the two classes. This careful rating method gives unusually high reliabilities, as shown in the tables. The range was from .42 to .90 with a mean of .60. These results show a fair consistency between the two sets of ratings for the different traits.

TABLE 2

RELIABILITY OF RATINGS: SELF-RATINGS CORRELATED WITH MEAN OF TWO ASSOCIATES' RATINGS

| Number of cases | Hysteroid-Psych. Deviate | Manic-Hypomania | Depression | Autistic-Schizoid | Paranoia |
|---|---|---|---|---|---|
| 40 | .56 | .74 | .42 | .56 | .54 |
| 56 | .82 | .55 | .90 | .55 | .42 |

TABLE 3

CORRELATIONS BETWEEN HUMM-WADSWORTH AND MEAN OF ALL RATINGS

| Number of cases | Hysteroid-Psych. Deviate | Manic-Hypomania | Depression | Autistic-Schizoid | Paranoia |
|---|---|---|---|---|---|
| 45 | −.28 | −.06 | .40 | .61 | — |
| 42 | .30 | .18 | .03 | .54 | .04 |
| 50 | .03 | .11 | .37 | .30 | .53 |
| 37 | .23 | .14 | .41 | .36 | .23 |

Table 3 gives the correlations between the Humm-Wadsworth scores and the mean of the three ratings for each of the five components for a total of 174 subjects. All but 2 of the 20 correlations were positive but low. The highest correlations were for the depres-

TABLE 4

CORRELATIONS BETWEEN MMPI AND MEAN OF ALL RATINGS

| Number of cases | Hysteroid-Psych. Deviate | Manic-Hypomania | Depression | Autistic-Schizoid | Paranoia |
|---|---|---|---|---|---|
| 42* | .28 | .23 | .26 | −.03 | −.12 |
| 50* | .05 | −.22 | .32 | .27 | .03 |
| 28* | .35 | −.29 | .49 | .09 | .27 |
| 39† | −.21 | .32 | .30 | .07 | −.14 |
| 38† | .43 | .10 | .48 | .15 | .19 |

*Group was administered the MMPI Group Form.
†Group was administered the MMPI Individual Form.

sion and autistic components. The mean correlation for these two components was .38.

The first part of Table 4 gives the correlations between the ratings and the group form of the MMPI for 120 subjects. Four out of 15 of these correlations were negative. The two components giving the highest correlation were psychopathic deviate and depression. The mean of these two was .30.

The last part of Table 4 gives the correlation of ratings with the individual MMPI for 77 subjects. The correlations were not consistently higher than for the group form. These results raise the question as to whether with normal subjects the individual form of the MMPI is more valid than the group form for the evaluation of personality. They do not seem to justify the additional time necessary to give the individual form of the test.

In summary it may be stated that insofar as carefully collected ratings with mean reliabilities of .60 can be used for validation, both the Humm-Wadsworth and the MMPI show fair validities for measuring certain personality components and not much validity for others. Each correlates considerably higher with ratings of these components than with each other. These facts may be stated in another way by saying that the seemingly common components of the two traits each measure different aspects of these traits.

These results should be interpreted, not as a criticism of the tests, but as suggesting the care and tentativeness that should be attached to any diagnosis based exclusively upon such personality test scores. The results also suggest that a new test might be constructed by combining certain parts of each of the present tests, giving a measure of higher validity for normal subjects than either the Humm-Wadsworth or the MMPI used separately.

*Received February 13, 1951.*

# THE RORSCHACH AND DESCRIPTIVE DIAGNOSIS

J. R. WITTENBORN[1]
YALE UNIVERSITY
AND
JULES D. HOLZBERG
CONNECTICUT STATE HOSPITAL

ALTHOUGH the Rorschach test has been used primarily to evaluate the structure of the individual's personality and his characteristic ways of coping with certain motives, it has also enjoyed considerable use as an aid in the descriptive diagnosis of mental hospital patients. The Rorschach test has been used as a diagnostic supplement in the sense that the Rorschach examiner has commonly stated the kinds of symptoms the patient is currently manifesting or is likely to manifest in the very near future.

The present report describes the relationships found between specific Rorschach scores and the kinds of symptoms that patients were known to manifest at the time the Rorschach test was given. The sample upon which the present analyses are based comprises 199 cases drawn from consecutive admissions at the Connecticut State Hospital at Middletown. The basis for selecting these 199 cases was twofold. All the patients were to be under 60 years of age, and none of the patients was to be suffering from a disorder which is known to be primarily neurological or toxic in its nature.

The validity of the various scores of the Rorschach test was evaluated by means of an employment of two criteria:

1. *Psychiatrists' diagnosis.* From the available sample the five most commonly occurring diagnostic groups were selected: paranoid schizophrenia; manic-depressive psychosis, manic phase; involutional psychosis; simple schizophrenia; and manic-depressive, depressed phase. Differences between the various diagnostic groups are sought for each of 39 different Rorschach scoring categories. Inasmuch as a psychiatric diagnosis is not a quantified attribute, the presence of relationship between a diagnosis and a Rorschach scoring category is sought by an application of the $\chi^2$ test. This is illustrated in Table 1, which shows that the manic-depressed patients in a manic phase produced *more* of the *CF* Rorschach determinant score than would be predicted from the general prevalence of the *CF* score among all the 75 patients included in the five diagnostic groups. Table 1 also shows that the manic-depressed patients in a depressed phase showed *less* of the *CF* score than would be expected for the total group of 75 patients falling in the five diagnostic groups in question. In Table 1, as in all the $\chi^2$ tests, the expected frequency of scores is based on a distribution of scores for the group of 75 patients resulting when the five diagnostic groups are combined. It should be noted that the first diagnostic group, paranoid schizophrenia, is the major contributor to the total and, as a consequence, the distribution of scores based on the total is biased in the sense that it will most resemble the distribution for the paranoid schizophrenic group. From some standpoints this is an undesirable basis for comparing the different diagnostic groups with respect to Rorschach performance; on the other hand, the present grouping is considered to be a typical one and represents the practical situation.

2. *Symptom-cluster scores.* The criteria for the second analysis are the patients' scores for each of nine different, empirically determined syndromes or symptom clusters [1, 3, 4]. The patient's score on each of these nine clusters is based on psychiatrists' ratings and provides a kind of a quantified multiple psychiatric diagnosis. The nine scores evaluate the patient in terms of most of the different aspects of behavior which are currently considered by psychiatrists to be important in a description of mental hospital patients. Due to its relative objectivity, standard nature, and quantified features, this type of description of the patient's current symtomatology

[1]The writers wish to acknowledge the collaboration of the psychiatric staff at the Connecticut State Hospital at Middletown and the support of Mark A. May, Director, Institute of Human Relations, Yale University.

should be more favorable for research purposes than the usual diagnosis of the psychiatrist. It should be pointed out that the symptom ratings were made during the first few weeks of the patient's hospitalization and at about the time the Rorschach test was administered.

There may be some readers who are inclined to question the relevance of a validation study of discrete Rorschach scoring categories. Inasmuch as isolated categories are not used in Rorschach practice, this procedure may at first glance seem unreasonable. In an earlier report [2], however, the writer has shown that no combination of scores such as the differences and ratios employed in scoring the Rorschach can bear a higher relationship with a criterion than an optimal weighting of the component scores. For this reason, it seems appropriate to study the validity of the component scores of the Rorschach test because if there is an evidence for the validity of these component scores, we may be optimistic about the validity of combining them for the prediction of different descriptive criteria. If, however, none or relatively few component scores have a validity for predicting the descriptive aspects of the patient's manifestations, the present data, at least, cannot be employed to support any optimism concerning the possibilities of using combinations of scores for predicting the descriptive status of the patient. It should be noted, however, that any validity or lack of validity revealed by the present sample cannot be generalized to outpatient samples, nor is it reasonable to suppose that any validity or lack of validity found for the descriptive use of the Rorschach test may be applied to other uses such as evaluations of the individual's motivational pattern. It should also be realized that there are many important features of the Rorschach test which are commonly used both for descriptive and dynamic evaluations which are not included in the present analyses, e.g., characteristics of the subject's approach, mode of response to the testing situation, and idiosyncratic or peculiar content.

The Rorschach scoring categories employed in the present analyses are: *W*, *DW*, *DdW*, *WS*, *D*, *Dd*, *DdD*, *Ds*, *Dds*, *R*, *F+*, *F*, *F—*, *M*, *FM*, *m*, *FC*, *CF*, *C*, *FC'*, *C'F*, *C'*, *Fc*, *cF*, *FV*, *FY*, *A*, *H*, *Hd*, *At*, Xray, Arch, Geo, Relig, eyes, *P*, *O*, color shock, shading shock.

## RESULTS

The results of both analyses 1 and 2 are rather discouraging from the standpoint of using the Rorschach scoring categories or any combination thereof in the prediction of the symptoms patients are likely to manifest.

TABLE 1

DISTRIBUTION OF *CF* RESPONSES FOR PATIENTS IN EACH OF THE DIAGNOSTIC GROUPS

| Diagnostic Group | *CF* Responses: None | *CF* Responses: One or More | Total |
|---|---|---|---|
| Paranoid Schizophrenia | 15(15) | 15(15) | 30 |
| Manic | 3( 7) | 10( 6) | 13 |
| Involutional | 5( 7) | 7( 5) | 12 |
| Simple Schizophrenia | 6( 5) | 4( 5) | 10 |
| Depressed | 10( 5) | 0( 5) | 10 |
| Total | 39 | 36 | 75 |

The parenthetical values are the frequencies expected from the hypothesis of uniform distribution.

Only 1 of the 39 $\chi^2$ tests required by analysis 1 resulted in statistically significant differences between diagnostic groups. These differences are for the *CF* Rorschach score (Table 1), and it seems that the presence or absence of the *CF* score may be a useful aid in distinguishing between patients who are in a manic or a depressed phase. Unfortunately, the distinctions between patients who are in a manic or depressed phase does not often require the employment of a supplementary device such as the Rorschach test. Other trends are suggested by analysis 1; these were significant at only the 10 per cent level. For example, there is a slight tendency for the manic-depressed patients in a manic phase to show the human movement response, whereas the human movement response is shown less often than would be anticipated for involutional patients. Similarly, there was some evidence that involutional patients will show more of the *Fc* responses than the other diagnostic groups included in analysis 1, whereas the manic depressed patients in a depressed phase tend to avoid the *Fc* response. The results show some slight tendency for the involutional patients to make a relatively prevalent use of human details. There may be some tendency for the paranoid schizophrenic patients to use

relatively more of the whole responses involving space than the other patients; there is also a slight tendency for the involutional patients to avoid the use of whole reponses which involve space. In general, these trends are congruent with our beliefs concerning the Rorschach performance of mental hospital patients and contribute nothing to our beliefs concerning the Rorschach test. On the other hand, in these analyses there is a conspicuous lack of statistically significant trends. This disappointing fact obviously cannot be used as a basis for saying that the Rorschach is an inappropriate device for descriptive diagnosis, but it is sufficient to warrant extreme caution and skepticism in using the aspects of the Rorschach test in question as aids in predicting the symptomatic behavior of patients, particularly the aspects of the patient's behavior which psychiatrists would employ in a descriptive diagnosis.

In order to satisfy the requirements of analysis 2, each of the 39 Rorschach scores was correlated with each of the nine symptom cluster scores. The correlation coefficients employed were either Pearson *r*'s or the tetrachoric correlations. The shapes of the distributions for the various Rorschach scores differ greatly, and it is improbable that the sampling distributions for the various correlation coefficients are highly similar. The correlation coefficients are based on a sample of 117 cases, but these correlation coefficients are doubtlessly less reliable than correlation coefficients based on normally distributed samples of this size. For lack of an alternative that is both desirable and feasible, correlation coefficients of .20 are arbitrarily taken as significant. Only 3 per cent of all the possible correlation coefficients met this arbitrary criterion for significance. Accordingly, it appears possible that the number of large correlation coefficients found in analysis 2 is not greater than would be expected by chance.

There is little justification for elaborate discussion of these correlation coefficients. Two of the Rorschach scores, achromatic color and the number of popular responses, show more than one correlation with the cluster scores. Achromatic color was correlated to the order of .23 and .27 with the manic cluster and the paranoid condition cluster, respectively. The number of popular responses was negatively correlated to the order of .20 with the excited cluster, negatively correlated to the order of .30 with the paranoid schizophrenic cluster, and negatively correlated .22 and .21 with the paranoid condition cluster and the hebephrenic or deteriorated schizophrenic cluster, respectively. The fact that the number of popular responses is negatively correlated with some of the psychiatric syndrome scores is plausible and corresponds with the common implications given popular Rorschach responses. It is also of interest to note that the color shock score is correlated to the order of .20 with the anxiety cluster. This is also congruent with the manner in which color shock is used as evidence for anxiety and other types of neurotic defects. The results of analysis 2, like those of analysis 1, are not encouraging for the use of Rorschach scores as aids for the description of the symptomatic manifestations of psychiatric patients.

## CONCLUSIONS

The present study, based on newly admitted patients to the Connecticut State Hospital, examines relationships between commonly employed Rorschach scores and a variety of descriptive aspects of the behavior of hospitalized functional psychotics. The present study comprises two parts. The first part compares the distributions of Rorschach scores for groups of patients with well-known psychiatric diagnoses. The second part correlates Rorschach scores with scores for empirically determined symptom clusters.

The present analyses cannot be taken as an evidence that descriptive aspects of the behavior of psychotic patients may be identified by or predicted by a use of Rorschach scores. Although the results of the present analyses must be limited to the use of Rorschach location and determinant scores and to the descriptive aspects of mental hospital patients' behavior, it is suggested that caution, if not skepticism, be employed in attempts to use Rorschach scores as an aid in descriptive diagnosis.

*Received February 16, 1951.*

## REFERENCES

1. WITTENBORN, J. R. A new procedure for evalu-

ating mental hospital patients. *J. consult. Psychol.*, 1950, **14**, 500-501.

2. WITTENBORN, J. R. An evaluation of the use of difference scores in prediction. *J. clin. Psychol.*, 1951, **7**, 108-111.
3. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, **15**, 290-302.
4. WITTENBORN, J. R., AND HOLZBERG, J. D. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, **15**, 372-380.

# IMPULSIVITY AS INDICATED BY RORSCHACH TEST FACTORS

RILEY W. GARDNER
UNIVERSITY OF KANSAS

MANY writers on projective techniques contend that major criticisms of the Rorschach test as an instrument for predicting overt behavior center around problems of validation of test factors. Schafer clarifies this argument considerably by pointing out that interpretations rather than scores are the proper units for research in projective tests and states that the actual need is for validation of these intepretations [9, p. 332]. In an experiment designed to test the supposed Rorschach indicators of intellectual control, Meyer Williams [10] would appear to have made a step in this direction. He found significant and reliable correlations between interpretations of a few selected test factors and a quantitative measure of intellectual control in an anxiety-arousing experimental situation.

The present study is intended to be a further step along this line. The general hypothesis is that rankings of subjects based upon common interpretations of Rorschach score factors applying to an impulsivity-inhibition continuum will correlate significantly with rankings of these subjects by raters from (1) long-term acquaintance with the subjects, and (2) brief behavior samples, in this case responses to the Rosenzweig P-F Test [8].

The definition of impulsivity used in this study was drawn from Rapaport's writings on the Rorschach test [5, pp. 238-241]. This definition is based upon a conception of personality which equates control with ability to delay response to external stimulation. If personality can be said to consist in large part of those mechanisms which are gradually interposed between basic impulses and the environment during the course of development, the strength and quality of these mechanisms will be indicated by the degree of control which can be observed in the person's response to externally determined situations. If affect can be considered, furthermore, a momentary derivative of original drives and these interposed mechanisms [cf. 5, p. 239], the person's way of handling affect, its quality, and its extent should be distinctive features of his response to any situation. The question investigated in this study is, then, whether or not Rorschach factors interpreted by four prominent writers on the test (Beck, Rapaport, and Klopfer and Kelley) as indicators of impulsivity allow us to predict that a person will also be "impulsive" in his everyday life.

## RATIONALE FOR THE INTERPRETATIONS OF TEST FACTORS USED IN THIS STUDY

A survey of the major works of these four writers on the Rorschach test [1, 3, 5] provided the experimenter with a group of interpretations generally agreed upon as in some way indicative of varying degrees of impulsivity.

*Color Responses.* Rapaport states that the pure color response (*C*) "represents either the extreme of impulsive and wild affectivity, or an abandonment of all control" [5, p. 242]. Beck summarizes his interpretation of the *C* response as follows: "The undiluted color reaction, *C*, is the test's equivalent of the uninhibited feeling experience" [1, p. 27]. Klopfer and Kelley formulate their interpretation in a similar way:

> Reactions like "blood" or "fire" given indiscriminately to any red or pink spot in all ten cards, or like "sky" or "water" to any trace of blue, reveal that the last remainders of a rational permeation have been extinguished in the emotional reactions represented by such responses [3, p. 286].

In the present study the *C* responses were used in three ways (see Table 1): in ratios representing (1) the relationship of *CF* (color-form) and *C* responses to *FC* (form-color) responses, (2) the proportion of *CF* and

*C* responses in the total record, and (3) the proportion of responses in the total record involving color. They were also included in the calculation of the Experience Balance for each subject (the sum of the movement over the sum of the color responses). Throughout this study color responses were weighted as follows: *C*, 1.5; *CF*, 1; *FC*, .5.

The four writers agree that the *CF* response is characteristic of a less impulsive reactivity than is the *C* response [cf. 1, p. 28]. The *FC* response is considered a still less impulsive and more socially conditioned one, which indicates the subject's ability to deal effectively with his affects and to modify them appropriately in reference to external reality.

*FC* responses were used in two ways in the present investigation: (1) in their relationship to *CF* and *C*, and (2) as one element in the total color responses as they relate to the total number of responses (*R*). The first use is based upon the assumption that the *FC* response indicates much more control of affectivity than does either the *CF* or *C*. The second is based upon the assumption that responses involving any amount of color determination are more representative of tendencies toward impulsive expression of affect than are responses determined solely by form qualities of the blots.

*The Experience Balance.* This ratio (*sum M*: *sum C*) is among the best known of all Rorschach test factors. Movement responses are traditionally associated with internal stimulation, in such forms as may result in fantasy activity, whereas the various color responses are associated with outgoing affective responses to environmental stimuli. Of this ratio Rapaport writes as follows:

> . . . taking the movement and color responses together, it is possible to obtain an estimate of the degree of coarctation or dilation of the subject. We have also emphasized that the "direction" of the qualitative wealth is indicated by the relative prevalence of movements over colors, or vice versa. This "direction" reflects the relative emphasis placed by an individual upon thinking and the delay of action it implies, or upon free affective expression and upon the corresponding modes of control of action [3, pp. 265–266].

Beck states that this ratio represents "the balance of introversive against extratensive trends, of fantasy living against outward expression of the feelings" [1, p. 60]. Klopfer and Kelley feel that color responses "reflect the reaction to 'stimuli from without'" in contrast to movement responses, which are resultants of "'promptings from within'" [3, p. 205]. In this study the usual ratio, *sum M*: *sum C*, was used to represent the balance between movement and color responses in the subjects' Rorschach records.

*Form Level of Color Responses.* It has long been recognized that the degree to which a subject is able to maintain a high level of form quality in responses involving color is an important indicator of the amount of control present. Our four writers are in agreement upon this point.

In an attempt to use this relationship between form level in form-color responses and form level in other responses, a ratio was used in this study between the $F + \%$ of the *FC* responses in each record and the $F + \%$ of all responses in which form is the principal determinant. The use of this ratio was in part determined by the fact that in the Rapaport scoring system, used throughout this study, *CF* as well as *C* responses are not scored with respect to form level. This may well have reduced the effectiveness of correlations such as those calculated from these Rorschach records.

*Form Responses.* If the pure color response is interpreted to be the response most indicative of impulsive, explosive outbursts of affectivity, the pure form response is generally considered the one most indicative of emphasis upon "intellectualizing" and the most effective control in avoiding the disturbing influence of affects in the handling of the stimulus material. Klopfer and Kelley state, for example, that:

> The concentration of a subject's interest on the purely "formal" qualities of the blot material seems to indicate the degree to which the subject is inclined to repress or control less "formal," more personal or spontaneous reactions to the stimulus material [3, p. 203].

Only the percentage of $F+$ responses was used to reflect highly organized form responses, without inclusion of responses scored $F\pm$ in the Rapaport scoring system. Additional factors were the $F\%$ (percentage of each total record consisting of pure form responses) and Rapaport's "*new F%*," which expresses

the proportion of all the responses in an individual record in which form is either the sole or the principal determinant.

*Other Factors Used in This Study.* In addition to form and color indicators, four others were employed. The first two, the percentage of responses listed by Rapaport as "popular" and the percentage of animal and animal detail responses, were used since they are generally interpreted to vary roughly with restriction in affective expression and with conformity to social norms. The third, the size of *R*, was used in an exploratory sense only, since it is the resultant of a mingling of factors, including many which are not the direct subject of this investigation. A variation of *R* was also used, on the basis of suggestions in Beck's writings — a ratio between the number of responses to the first seven cards and responses to the last three (none of the former being entirely colored).

Thus, an important segment of the literature on the Rorschach test yields rather consistent interpretations of various test factors relating to an impulsivity-inhibition continuum. It is the purpose of this study to show, in a very limited way, whether these interpretations agree with actual observations and ratings of the overt behavior of individual subjects. It is granted that each writer quoted may be speaking at one time or another of *potentialities* for action, rather than overt behavioral phenomena. In general, however, their interpretations are intended to allow predictions of the reactions of individuals to life situations, and it is in this way that they will be considered here.

## HYPOTHESES

Preliminary hypotheses were: (1) that significant intercorrelations would be found among the rankings of the subjects by the three raters assigned to each behavior sample (long-term acquaintance and responses to the Rosenzweig P-F Test), and (2) that a similar correlation would exist between the average rankings of the subjects by the two groups of raters.

Specific hypotheses concerning the interpretations of the Rorschach test factors appearing in Table 1 were based upon pre-experimental judgments of the degree to which rankings of the subjects by means of these interpretations would correlate with the two criteria (the average ranking by each of the two groups of raters). These judgments were based largely upon the degree of confidence which the four writers place in individual interpretations and the consistency with which these interpretations appear in the writings. Thus, it was hypothesized that rankings based upon interpretations of the factors listed under A in Table 1 would show significant correlations with the two criteria. It was assumed that rankings based upon the factors under B would show lower correlations with these independent rankings of the subjects, but would nevertheless give some valuable indications in respect to impulsivity. It was hypothesized that rankings based upon the factors under C, in spite of their mention in the literature as fair, and sometimes good, indicators of impulsivity, would not show significant correlations with the criteria.

TABLE 1

RANK-ORDER CORRELATIONS BETWEEN SUBJECTS' VALUES ON 12 TEST FACTORS AND ESTIMATES OF THEIR IMPULSIVITY BY TWO GROUPS OF RATERS

| Rorschach Scoring Category | Estimates of Impulsivity Based on | |
|---|---|---|
| | Long-Term Acquaintance | P-F Test Responses |
| A. High correlations predicted | | |
| *CF + C* (including *C/F*) : *FC* | .879 | .815 |
| *CF + C* : *R* | .865 | .680 |
| *FC + CF + C* : *R* | .788 | .609 |
| *Sum M* : *Sum C* | .705 | .427 |
| *F + %* | .549 | .418 |
| B. Fair to good correlations predicted | | |
| *F + %* of *FC* responses : *F + %* of all form-primary responses | −.191 | −.491 |
| *F%* | .342 | .261 |
| *New F%* | .277 | −.006 |
| *Popular* : *R* | .670 | .855 |
| *A + Ad* : *R* | .518 | .636 |
| C. Low correlations predicted | | |
| Responses to Cards I-VII: VIII-X | .349 | .179 |
| *R* | .382 | .482 |

## PROCEDURE

Subjects were eight men and two women.

The only specific criteria applied to them were that they be university undergraduate or graduate students (an attempt to gain some control over the level of intelligence) and that they be known personally to all three "long-term" raters for a period of at least eight months immediately prior to the investigation. They ranged in age from 21 to 27.

The three long-term raters were graduate students in clinical psychology who were selected because of their mutual acquaintance with all ten subjects and because it was felt that they were qualified to rank these subjects on an impulsivity-inhibition continuum. They used Rapaport's statements about impulsivity [5] as the basis for their independent rankings. The three raters who ranked the subjects by use of the Rosenzweig P-F Test records were also graduate students in clinical psychology. They were not acquainted with the subjects and were given no clues as to their identity. Their ratings were made independently.

Each subject was administered two tests by the writer: (1) the Rorschach [7], in accordance with the directions for administration offered by Rapaport [5], and (2) the Rosenzweig P-F Test [8]. The latter test was used because it seems to provide, in a very brief period, a good sample of behavior with respect to the impulsivity-inhibition continuum. The Rorschach records were scored according to Rapaport's suggestions and then rescored, with a careful cross-check being made to insure uniformity. The cross-check was used to offset any influence the examiner's personal acquaintance with the subjects might have had on his scoring of their responses. The Rosenzweig P-F Test records were given to each of the three raters allotted to them in their original, unscored form and without any identifying marks. They were arranged in an order based upon the previous rankings of the subjects by the long-term raters in such a way as to avoid influence of order upon the new rankings.

Values for the twelve factors listed in Table 1 were then extracted from each Rorschach record, and the ten subjects were ranked according to the interpretation of each factor, in the order of impulsivity indicated by the relative strength of that factor. Spearman rank-order correlations were computed for the relationship between each of the twelve rankings and the two criteria. The significance of these correlations was determined on a scale derived from the formula for $t$ in small samples, described by Lindquist [4, p. 211] and Garrett [2, p. 346], which does not assume normality of the sampling distribution.

## RESULTS

The subjects displayed a wide range of responsiveness to the series of ink blots and to various determinants in the blots. Total number of responses ($R$) varied from 15 to 105. The Experience Balances of the subjects were: 4/0, 5/1, 10/3, 18/6.5, 5/6, 11/17.5, 6/16.5, 1/4, 2.5/10, and 1/5. The $F\%$ (percentage of $R$ in which form was the sole determinant) ranged from 46 to 73.

Intercorrelations between rankings of the subjects by the three long-term raters were .600, .769, and .346. Inquiry revealed that the low correlation, .346, between two of the long-term raters resulted largely from a marked difference of opinion about one subject. Between the three raters ranking the subjects by means of the Rosenzweig test, the correlations were .830, .855, and .721. The correlation between the average rankings by the two groups of raters was .815. (In this study the .01 level is .765, the .05 level is .632.)

In the first hypothesis about interpretations of the test factors it was stated that the factors under A in Table 1 would have high correlations with the average ranking by each set of raters. This appears to be borne out for the most part in the case of the long-term raters, but to be less so for the raters of the Rosenzweig test records. The factors listed under the second hypothesis seem not to fit our expectations, with the possible exception of the *Popular*:$R$ ratio and the $A + Ad$:$R$ ratio. The former slightly exceeds the .05 level in relationship to the long-term raters and exceeds the .01 level with the raters of the Rosenzweig test records. The latter achieves the .05 level only in relationship to the raters of the shorter behavior sample. As was expected, none of the factors listed under the third hypothesis approached significance. Thus, the first and third hypotheses were fairly well substantiated by the results; the second was

not. Among those factors which show significant correlations, as might be expected, the long-term ratings appear to be generally more satisfactory than the ratings of brief behavior samples.

## CONCLUSIONS

Two general conclusions arise from this study: (1) this type of investigation seems to have some value in determining the validity of various interpretations of Rorschach factors; and (2) there is serious need for careful determination of the predictive powers of commonly accepted interpretations. Although in this limited study the Rorschach test seems to contain several factors which have predictive value for the overt behavior of individual subjects, other factors interpreted with similar confidence do not correlate significantly with behavioral criteria. It is true that in clinical practice these factors are not interpreted singly, but in clusters. Much of the test's usefulness, however, depends ultimately upon the validity of interpretations of individual factors which make up these clusters.

As might be anticipated, rankings based upon considerable day-to-day acquaintance with subjects correlate better with test interpretations than do rankings based upon a brief sample of behavior such as the Rosenzweig P-F Test. The correlation of .815 between the average rankings by the two groups of raters suggests, however, that the P-F Test when used in this way offers a surprisingly rich sample of behavior with respect to an impulsivity-inhibition continuum.

More specifically, those factors appear to be rather good for predictive purposes in which quantitative aspects of various color responses are placed in ratios with *R*, or in which the more strictly color responses are placed in ratio with responses determined by both form and color, with form primary. The Experience Balance and the $F+\%$ may also be rather good indicators as they are generally interpreted. Other promising interpretations are those concerning the percentages of popular and animal responses, their size being said to vary inversely with the degree of impulsivity. Such factors, however, as $F+\%$ of $FC$ responses; $F+\%$ of all form-primary responses; $F\%$; *new* $F\%$; $R$; and a variation of $R$, are found to be very poor indicators for the relative degrees of impulsivity characterizing the subjects of this study.

*Received February 24, 1951.*

## REFERENCES

1. BECK, S. J. *Rorschach's test.* New York: Grune & Stratton, 1946. Vol. II.
2. GARRETT, H. E. *Statistics in psychology and education.* New York: Longmans, Green, 1947.
3. KLOPFER, B., AND KELLEY, D. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
4. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
5. RAPAPORT, D., GILL, M. M., AND SCHAFER, R. *Diagnostic psychological testing.* Chicago: Yearbook Publishers, 1946. Vol. II.
6. RORSCHACH, H. *Psychodiagnostics.* Berne: Hans Huber, 1949.
7. RORSCHACH, H. *Psychodiagnostics.* Plates. Berne: Hans Huber.
8. ROSENZWEIG, S. The picture-association method and its application in a study of reactions to frustration. *J. Personality,* 1945, **14,** 3-23.
9. SCHAFER, R. Psychological tests in clinical research. *J. consult. Psychol.,* 1949, **13,** 328-334.
10. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.,* 1947, **11,** 21-29.

# EXPERIMENTALLY INDUCED FAKING OF TAT RESPONSES [1, 2, 3]

EDITH A. WEISSKOPF
AND
JORGE J. DIEPPA
PURDUE UNIVERSITY

WHEN an individual is given a projective test, the outcome is frequently of vital importance for his future. Whether or not he is selected for a position may depend, in part, on his performance. If he wants the position, he may make a conscious effort to give responses which make him appear the kind of person most suited for the particular occupation. On the other hand, a person who wishes to be discharged from the armed forces may attempt to respond in such a way as to appear a bad prospect for adjustment to military service. It is important therefore, that projective techniques should be resistant to such distortive malingering. The work of Fosberg [3, 4, 5] seems to indicate that the Rorschach fulfills this requirement. On the other hand Carp and Shavzin [2] maintain that Rorschach responses can be faked. Similarly, a study by Meltzoff [6] with a sentence completion test indicates that this test may be subject to distortion.

A test such as the TAT can be expected to be subject to faking to a higher degree than a test such as the Rorschach. This expectation is based on the impression that subjects are more aware of the principles of interpretation with the TAT than with the Rorschach.

It is the purpose of the present study to test the following hypothesis: Subjects can influence the evaluation of their personalities based on TAT stories by trying to make a "good" or "bad" impression on the examiner. An examination of some of the methods used by the subjects to create a good or bad impression is a further objective of this paper.

## SUBJECTS

The subjects were 24 hospitalized male veterans, diagnosed as psychoneurotic. The reason for the use of psychoneurotic instead of so-called "normal" subjects was their availability to the junior author in his capacity as Psychologist Intern at a Veterans Hospital. The subjects' ages ranged from 23 to 48 years, with a mean age of 34.1 years.

## PROCEDURE

Three cards of the TAT series were administered to each subject. The cards were: No. *1*, a boy with a violin; No. *8BM*, a boy in the foreground, two men performing a surgical operation in the background; No. *13MF*, a man standing, a woman with bare breasts lying in bed.

The rationale for the selection of these pictures was as follows: (a) It was attempted to select pictures which elicit a relatively high amount of projection. According to a previous study by Weisskopf [7] the three above-mentioned pictures are among the ten pictures in the TAT series of twenty which elicit the highest amount of projection. (b) It was attempted to select pictures which elicit material on a variety of adjustment areas. The above-mentioned pictures usually contribute to the

[1]Published with the approval of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or conclusions drawn.

[2]The paper is based on a thesis presented to the Graduate School of Purdue University, Lafayette, Indiana, by the junior author in partial fulfillment of the requirements for an M. S. degree in psychology. The study was directed by the senior author.

[3]This study was reported at the annual meeting of the Midwestern Psychological Association, Chicago, April 1951.

diagnosis in the areas of level of aspiration (Cards *1, 8BM*), relationship to the parents (Cards *1, 8BM*), aggression and sadism (Cards *8BM, 13MF*), heterosexual impulses (Card *13 MF*), etc. (c) It was attempted to select pictures which would elicit stories most likely to be distorted when the subject tries to make a "good" or "bad" impression. This would be expected with pictures eliciting stories on such personality characteristics as level of aspiration or aggression which are strongly regulated by social mores, either through encouragement or through taboos.

It may be objected that the authors "loaded their dice" by selecting pictures which can, more than others, be expected to lead to corroboration of the hypothesis. However, such a procedure seems permissible if, in case of positive results, it is kept in mind that the lack of "foolproofness" has only been demonstrated for three pictures and that other TAT cards may be more resistant to faking.

Three sets of instructions were used in administering the pictures. One of these consisted of the traditional instructions as suggested by Murray. This set will be referred to hereafter as the Neutral instructions.

Another set of instructions contained the following suggestion: "This time I want you to tell your stories in such a way that, when I hear them, I will receive the *best* possible impression of you." These instructions will be referred to hereafter as the Best instructions.

The third set of instructions was formulated as follows: "This time I want you to tell your stories in such a way that when I hear them I will receive the *worst* possible impression of you." These instructions will be called the Worst instructions in the present discussion.

All subjects were first tested with the three pictures under Neutral instructions. Then the following explanation was given: "This test can give information on what kind of a person you are. But you can fake your stories so that I may receive a good impression or a bad impression of you according to your wishes."

After this explanation, the three pictures were given to one half of the group of subjects under the Best instructions first and under the Worst instructions second. The pictures were administered to the other half of the group under the Worst instructions first and the Best instructions second. The Neutral instructions were administered first to each subject because it was assumed that it would be difficult to obtain a truly spontaneous "neutral" story after the subject had been exposed to the Best or Worst instructions.

In the following discussion, three stories, all obtained from the same subject by the same set of instructions, will be referred to as a *protocol.*

The protocols were distributed among three judges, trained and experienced in the diagnostic use of the TAT. Each judge received 24 protocols, that is, *one* protocol from each subject. The protocols received by each judge included an equal number obtained by the Neutral, Best, and Worst instructions.

Moreover, each judge received the same number of Best protocols produced prior to the Worst protocols as Best protocols produced after the Worst protocols. The protocols given to each judge were arranged in random order and then numbered from No. 1 to No. 24. The order was different for the three judges.

The judges were asked to rate each subject, in the order indicated by the number of the protocols, as very high, high, average, low, or very low, with respect to the following characteristics:

1. Over-all adjustment
2. a. Total hostility (repressed and conscious)
   b. Conscious control of hostility
   c. Repression of hostility
   d. Guilt about hostility
3. a. Total sexual impulse (repressed and conscious)
   b. Conscious control of sexual impulse
   c. Repression of sexual impulse
   d. Guilt about sexual impulse
4. Level of aspiration
5. Eagerness to conform
6. Tendency to help others
7. Tendency to be helped by others
8. Intelligence
9. Spontaneity

It was stated in the instructions that ratings could be omitted if no evidence was found in the protocols.

The judges, when rating the subjects, did not know that part of the protocols had been obtained by instructions different from the

traditional ones. However, each judge received a sealed envelope with the instructions that it should not be opened until after the ratings had been completed. The following text was enclosed in the envelope: "The protocols you have just read were obtained by the following three different instructions: (A description of the three types of instructions was given.) Go over the protocols again and mark each one Neutral, Best, or Worst according to the instructions which you think produced the story."

## RESULTS

In order to be able to manipulate statistically the qualitative ratings on the rating sheets, numerical values were assigned to these ratings as follows:

| | |
|---|---|
| Very high - - | 5 |
| High - - - - | 4 |
| Average - - - | 3 |
| Low - - - - | 2 |
| Very low - - - | 1 |

Analysis of variance was used to evaluate the data. Three variables were considered: instructions, judges, and subjects. Only the results relevant to the hypothesis will be reported here, namely, those pertaining to the variations in instructions.

Variations in instructions produced a difference statistically significant beyond the 1 per cent level in the ratings of overall adjustment, total hostility, and eagerness to conform, and a difference significant between the 5 per cent and 1 per cent level in the ratings of total sexual impulses and spontaneity. The ratings on the other questions were not significantly affected by variations in instruction.

It appears that the subjects were rated as less well adjusted, more hostile, less eager to conform, of stronger sexual impulse and more spontaneous when the ratings were based on the Worst stories than when they were based on the Best and Neutral stories. There was a tendency for the Neutral ratings to be between the Best and Worst ratings, but considerably closer to the Best ratings.

The latter finding is consistent with the results of the judges' attempts to identify the Neutral, Best, and Worst protocols. The judges were able to label correctly 79 per cent of the Worst protocols, but only 58 per cent of the Best protocols and 58 per cent of the Neutral protocols. The differences between the per cent of correct identifications of Worst protocols and of Best or Neutral protocols is significant beyond the 1 per cent level. Thus, the Worst protocols were characterized by distinctive features to a higher degree than the Best and Neutral protocols.

The following quotations from the protocols of a subject illustrate qualitatively some of the results of this study, especially the increased aggression and sex impulse elicited by the Worst instructions. The example also illustrates that the Neutral stories tend to be more similar to the Best stories than to the Worst stories.

The subject told the following Neutral story about Picture *13MF*:

> Well, it looks to me like this woman is sick or something and that is her husband there, well, looks awful downcast because of her condition. That is about all I can see out of it. Maybe she will get better, maybe she won't. That is about all I can make of it. That's all.

Here is the Best story, told by the same subject to the same card:

> I would say that this woman had been sick with some illness for, — oh, like pneumonia or something like that. Has been between life and death for a long time and this man has been doctoring her or perhaps the books on the table could represent a, — one could represent a Bible. He could have been reading or administering to her. But finally she has come out of a coma or something and he knows she is going to be all right and is weary from watching over her face so long and now is either thankful, or praying, or something. Outcome is like it is and she is going to get better.

Finally, the Worst story:

> Well, let's see. Well, the only thing wrong is the man's hand over his face. He would be leaning over the woman and you would get the impression that this woman is a prostitute. The woman seems to be dead and the man has killed her and he is sorry for his deed. He killed her and now is sorry that he has done it. That would be all I can think of.

Another subject carries the desire to appear "asexual" when trying to make a good impression to the extreme. Here is his Best story about Picture *13MF*:

> This is a story of a great love of a young lady. Her beauty and enchantment this young fellow has admired since he was a boy. Having been reared

together, his love grew for her as they grew older. She in return returned this love although retaining her virginity. They waited long for the day of their parents' consent to be married. Their life was moral, decent, both members of the church helping those in the community who were less fortunate. Oh! the world looks so good in the future of these young people! They are married finally. They have a church wedding, friends, mothers and fathers attended. Having a wonderful time. They leave church to go on their honeymoon. This young fellow who has been shy although he possessed great love for the lady, when they went to bed he found it was impossible to go to bed because he felt that something was being destroyed. He is immature, unable to achieve the happiness of sexual desire which are normal for a man and wife. Their life has disappeared. He realizes that they are headed for the rocks.

### QUALITATIVE COMMENTS

A qualitative study of the protocols throws some light upon the mechanisms which the subjects use to make a favorable or unfavorable impression.

A comparison of the Best or Neutral protocols with the Worst protocols suggests the tentative interpretation that many subjects attempt to appear socially acceptable in the Best or Neutral stories by introducing defenses against the socially unacceptable impulses shown in the Worst stories. The defenses used are such well-known mechanisms as sublimation, projection, reaction-formation, rationalization, displacement, etc.

One subject, for example, tells the following Worst story to picture *8BM*:

Well, I would say that this was a young man given in two different ages. Almost like a dream. He has dreamt for a long time to bring the ending of another person, possibly his own father and that is what he has been doing for some time. He has been carrying this dream for some time and did not fulfill his wish until late in life when his father was old and decrepit. He is being helped by a friend, on the perverted side himself, who is only there to witness the pain and anguish of the man. I see in the future that he will be able to get along with society and the law but he will suffer himself because of his conscience in later life. I see him only as a murderer.

The Best story, told by the same subject to the same picture, runs as follows:

Well, I see this boy is dreaming and he has just witnessed for the first time—his father who is a doctor has allowed him to witness an operation. I see him dreaming and hoping that in the future he will have the hand of his father that he may be able to prolong life as he has seen it up to this time. He may not have any appreciation for his father until he has seen him now to know the tribulations while the patient is under the knife. The operation may prove unsuccessful and the boy became confused as to whether he will be able to withstand the shock of failure that may cause death; whether the reward of lengthening the life of some will be worth the pain of failure when you lose a case.

A comparison between the two stories suggests the following interpretations:

The crude sadistic activity in the Worst story is sublimated into surgical activity in the Best story. The hatred against the father in the Worst story is not apparent in the Best story. Instead, the hero desires to imitate the father, to make the father's ideals and skills his own. This mechanism is known as *identification*. Thus, the transition from the Worst to the Best story occurs through *sublimation* and *identification*. This transition resembles the socialization process of individuals: the antagonism against the parent of the same sex is being resolved by *identification* with parental ideas, and unacceptable impulses are directed into constructive channels.

The perception of Picture *8BM* as an aggressive scene under the Worst instructions and as a surgical operation under the Best instructions appears frequently in the protocols of this study.

Another subject tells a Worst story about Picture *8BM* to the effect that the boy in the foreground shot the person who is being operated on; the boy hopes this hated person will die during the operation. In the Best story the boy daydreams that he is a surgeon who performs an operation. In the Neutral story the boy in the foreground also daydreams of the operation in the background; this time, however, he is not the surgeon, but he fears that such an operation might be performed on himself by other people. Again, as with the previous subject, the Best story can be interpreted as a sublimated modification of the Worst story. The Neutral story, on the other hand, may be a *projective* modification of the Worst story, since the hero projects his aggressive tendencies upon other people and fears their retaliation for his own hostility.

In many cases the Best or Neutral story can be conceived as a *reaction-formation* against the antisocial impulses disclosed in the Worst story.

Thus, a subject tells the following Worst story to Picture 1:

This is a young lad who has stolen a package that he found in the hallway of an apartment. This package later turned out to be a Stradivarius violin and several musical scores which were in the case. He later, not knowing the value of the violin, took a hammer and smashed it to pieces. That hurts me when I say that the Stradivarius was smashed to pieces.

In the Best story of the same subject the hero shows extremely positive rather than extremely negative feelings towards the violin:

This boy is about 10 years of age. He has been a child prodigy since the age of four. He has just been given a violin by his old teacher, the violin being a Stradivarius. Although this lad is only ten he fully appreciated the value of such a beautiful instrument of music and he resolved never to let this violin part from him and also that he would become the world's foremost violinist which he later did.

The same kind of reaction-formation can be found in the Neutral story:

This is a young musical prodigy who is probably a master of the violin, and who has just laid his violin down across his music score and is gazing at the violin with adoration or reverence. The outcome of this, he will continue being a musician probably ending third fiddle on some symphony.

In the following illustration a comparison between Neutral and Best story suggests the use of *displacement* in addition to reaction-formation. The Neutral story to Picture *8BM* reveals overt aggression against a parent:

I think the story in the picture happened around the 1890's because the tie that this gentleman has on, the one holding the lantern, and the collar of his shirt. I should say that the operation is being performed because of a wound and this wound has been inflicted. I believe it is a mother by looking at the little breast. I cannot tell. I imagine that this young fellow, the son, has had an unhappy home life and has brought about family difficulties of some sort and this young fellow he mulled over the proposition of what happened at home without thinking, I should say, took that gun and decided to murder father or mother. His attempt was unsuccessful because they took the parent to the doctor who performed an operation, who is going to save the life of the parent. He has a look of regret and is wondering what the outcome of the operation will be because in turn it will determine his future life and his sorrow has caused him to realize that he should not have acted as he did. At the end the parent returned to normal health but there will always be doubt in this fellow's mind because of what he has done although the parent is willing to forgive. This is a case of another frustrated life.

In the Best story, the hero is strongly devoted to his mother:

Well, unhappiness has been brought to the life of this young fellow. We find that his mother has been shot and he realizes, and knowing the person that committed the crime but not knowing the reason, wonders what action he should take. If his mother could get well and would tell him why, it would give him ground to decide what he should do. He has made up his mind to shoot the person that has shot his mother if she is in the right. He is worried. "Oh, if she would only get home. If she would get well and tell me." He is praying that God will bring his mother around. We hope that this young fellow's prayers are answered. Not for the reason that he in turn can take this drastic action against the assaulter of his mother but that he will have the love of his mother. Here he is at twelve without a mother to keep him through the years of adulthood. She will come out well and will help him until he is twenty-one and can marry a girl that understands him.

Here the subject's concern for his mother can be interpreted as a reaction-formation against his aggression. However, the hero also plans to act out his aggression in an overt manner, but this time, it is displaced from his mother to a criminal. Thus, the hero is in a similar situation to that of a censor of pornographic art, who can satisfy his voyeuristic tendencies by fighting voyeurism.

In many cases antisocial impulses which are overtly expressed in the Worst story appear disguised through *rationalization* in the Best or Neutral story. For example, the hero of the Worst story may hurt or kill an enemy intentionally, while the same enemy may be killed by the hero accidentally in the Best or Neutral story.

The following stories may serve as an illustration. In the Neutral story to Picture *8BM* the subject does not specify whether the hero's father was shot accidentally or intentionally.

Well, the impression I get of a youngster that probably went hunting with his father and shot him and has a dream that they are operating on the father. Probably worried as to the outcome.

In the Best story an accident is the cause of the shooting:

Looks like a young man and father have been hunting and he accidentally shot his father. Has vision on his mind or dream wondering as to outcome of the operation being performed on the father. Whether it is going to be good or bad.

The Worst story, finally, shows overt aggression. Here, the shooting is done intentionally:

Could be that the boy intentionally shot his father and maybe dreams that if the bullet doesn't do the job right, maybe the doctor's knife will slip. That is supposing it would be a gun in there, is it?

The fact that not only the Best stories but also the Neutral stories appear to be censored modifications of the Worst stories lends support to Balken's and Wyatt's statement that stories of the Thematic Apperception type reflect mainly phantasies which are somewhat acceptable to the ego [1, 8].

## SUMMARY AND CONCLUSIONS

1. Subjects can, when trying to make a good or a bad impression, successfully influence the diagnosis of their personalities made by experienced TAT interpreters on the basis of their responses to three TAT pictures.

2. When trying to make a bad impression, the subjects tended to change their Neutral records to a higher degree than when trying to make a good impression. From this observation the tentative conclusion can be drawn that individuals taking the TAT may be attempting to make a good impression on the examiner, even if not explicitly required to do so by the instructions.

3. The Best and Neutral stories can be interpreted as reflecting defenses such as sublimation, projection, reaction-formation, and rationalization against the antisocial impulses expressed in the Worst stories.

## SUGGESTIONS FOR FURTHER RESEARCH

1. Since the manner in which people attempt to make a good or bad impression is dependent on the moral code of the group in which they live, the results of a study such as the present one can be expected to vary with the cultural group from which the subjects are selected. In the present study the subjects attempted to make a good impression by appearing less hostile, less spontaneous, less sexual, and more conforming than they actually are. In a less conventional group, for example, attempts to make a good or bad impression may be carried out in an entirely different manner. Thus, the technique used in this study may help to obtain information about the concepts of morality and respectability in specific groups.

2. From a qualitative inspection of the records it seems probable that a comparison of an individual's Neutral, Best, and Worst protocol would be of great diagnostic value. Such a comparison might be more informative regarding the individual's impulses, aspirations, and defenses than the customary inspection of neutral records. Thus, the methods used in the present study might be developed into a valuable diagnostic tool, which could supplement or replace the traditional TAT.

*Received August 6, 1951.*
*Early publication.*

## REFERENCES

1. BALKEN, EVA R. AND VANDER VEER, A. H. The clinical application of a test of imagination to neurotic children. *Amer J. Orthopsychiatry*, 1942, **12**, 68-81.
2. CARP, A. L., AND SHAVZIN, A. R. The susceptibility to falsification of the Rorschach psychodiagnostic technique. *J. consult. Psychol.*, 1950, **14**, 230-233.
3. FOSBERG, I. A. An experimental study of the reliability of the Rorschach psychodiagnostic technique. *Rorschach Res. Exch.*, 1941, **5**, 72-84.
4. FOSBERG, I. A. How do subjects attempt to fake results on the Rorschach test? *Rorschach Res. Exch.*, 1943, **7**, 119-121.
5. FOSBERG, I. A. Rorschach reactions under varied instructions. *Rorschach Res. Exch.*, 1938, **3**, 12-31.
6. MELTZOFF, J. The effect of mental set and item structure upon response to a projective test. *Amer. Psychologist*, 1950, **5**, 326. (Abstract).
7. WEISSKOPF, EDITH A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
8. WYATT, F. The interpretation of the Thematic Apperception Test. *Rorschach Res. Exch.*, 1947, **11**, 21-25.

# PUPIL FUNCTIONING ON THE STANFORD-BINET AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

JUDITH I. KRUGMAN, JOSEPH JUSTMAN, J. WAYNE WRIGHTSTONE
BUREAU OF EDUCATIONAL RESEARCH
AND
MORRIS KRUGMAN[1]
BUREAU OF EDUCATIONAL AND VOCATIONAL GUIDANCE, BOARD OF EDUCATION, NEW YORK, N. Y.

FOR years, psychologists in the New York City Board of Education, like those throughout the country, have used the Revised Stanford-Binet as the basic technique in measuring general intelligence. Although they have found it a most useful instrument, they have felt that it is very heavily weighted with verbal items and has yielded an inadequate measure for many children from culturally limited homes. Further, the test has not lent itself too well to diagnostic analysis of abilities because of the diverse types of items at different age levels.

When the Wechsler Intelligence Scale for Children was constructed, the psychologists were eager to try it out, for it seemed to answer many of their criticisms of the Binet. It included items to measure similar abilities at all age levels, permitted more opportunity for diagnostic analysis, gave a performance rating as well as a verbal and full-scale rating, and had material that was appealing to children and simple for the examiner to administer.

Before accepting an instrument like the WISC to take the place of the Binet, however, the psychologists felt that not only would it be necessary to try it out to substantiate their clinical impressions of its values, but they would have to study the results in comparison with the Binet results. The latter had been used as the measuring rod in making placements for special classes for children with retarded mental development in accordance with the provisions of a state law.

Dr. Morris Krugman, who was at that time chief psychologist in the Bureau of Child Guidance, and Dr. J. Wayne Wrightstone, Director of the Bureau of Educational Research, arranged with Dr. David Wechsler to administer the WISC for standardization purposes. Dr. Wechsler was to use those cases which fitted into his standardization scheme and to return the results of these and other cases so that the findings for New York City children could be studied and could be compared with the Stanford-Binet results.

The testing was to be undertaken by psychologists of the Bureau of Child Guidance and the Bureau of Educational Research, and the results were to be treated statistically by the Bureau of Educational Research.

## PROCEDURE

Schools were selected to give representation to some of the varied neighborhoods, socio-economic strata, and ethnic groups which make up New York City. In all, 18 schools in the five boroughs were used. The average IQ's on city-wide tests given in the sixth or seventh grades in these schools ranged from 88 to 118 on the Pintner Intermediate General Ability Test, with a median IQ of 100 for the 18 schools. Children were tested at each age level from 5½ years to 15½ years, if their ages

[1]The writers would like to acknowledge the assistance of the following psychologists who participated in the testing program: Nathan Altrowitz, Arthur Berger, Myra Z. Goodman, Louis Granich, Janet Griswold, Henry Hansburg, Georgiana T. January, Marion Jenkins, Judith I. Krugman, Dorothy B. Lee, Louise B. Lowenstein, Edward Moses, Esther Mullen, Josephine McGurk, Bertha Nelson, Bernice Plotkin, Morris Speevack, Lillian Tow, Thelma Voorhis, Hazel Wertman, and Margaret Williamson. Mr. Jacob Abramson, of the staff of the Bureau of Educational Research, contributed materially to the analysis of the data.

were within 1½ months of the half-year. An equal number of boys and girls were tested with both the WISC and Form L of the Revised Stanford-Binet. The order of presentation of these tests was alternated, the WISC administered first in half the cases, the Binet first in the other cases. Wherever possible, the tests were given on successive days. In a few instances children had had Binet examinations several months earlier. If these were given less than six months previously, they were included.

In the 1946-1947 school year 21 psychologists (19 from the Bureau of Child Guidance and 2 from the Bureau of Educational Research) administered the tests to children in the 18 schools. The findings reported here are based upon the 332 records returned (166 boys, 166 girls) for whom both WISC and Binet results were available. Of these, 171 records had been used by Wechsler as part of his standardization population [1].

TABLE 1

CORRELATION BETWEEN STANFORD-BINET AND WISC VERBAL, PERFORMANCE, AND FULL SCALES

| Age Level | N | S-B vs. WISC Verbal | S-B vs. WISC Performance | S-B vs. WISC Full Scale |
|---|---|---|---|---|
| 5 | 38 | .668 | .763 | .896 |
| 6 | 38 | .726 | .741 | .821 |
| 7 | 43 | .642 | .486 | .733 |
| 8 | 44 | .783 | .574 | .823 |
| 9 | 31 | .834 | .790 | .873 |
| 10 | 29 | .880 | .538 | .856 |
| 11 | 37 | .691 | .532 | .757 |
| 12 | 22 | .697 | .541 | .923 |
| 13 | 30 | .715 | .669 | .818 |
| 14–15 | 20 | .884 | .608 | .832 |
| Total | 332 | .739 | .644 | .817 |

## THE FINDINGS

As a first step in the treatment of the obtained data, product-moment correlations were

TABLE 2

MEAN IQ'S AND SIGNIFICANCE OF MEAN DIFFERENCES OF S-B AND WISC FULL SCALE IQ'S AT EACH AGE LEVEL AND FOR TOTAL GROUP

| Age Level | *N* | *Mean* | *SD* | $SE_M$ | *D* | $SE_D$ | *t* | *P* |
|---|---|---|---|---|---|---|---|---|
| 5 Binet | 38 | 107.26 | 14.90 | 2.451 | 11.18 | 1.09 | 10.257 | $P < .01$ |
| WISC | 38 | 96.08 | 13.20 | 2.171 | | | | |
| 6 Binet | 38 | 111.87 | 14.30 | 2.352 | 10.00 | 1.38 | 7.246 | $P < .01$ |
| WISC | 38 | 101.87 | 13.55 | 2.228 | | | | |
| 7 Binet | 43 | 105.02 | 11.60 | 1.790 | 3.95 | 1.25 | 3.160 | $P < .01$ |
| WISC | 43 | 101.07 | 10.30 | 1.589 | | | | |
| 8 Binet | 44 | 110.86 | 15.20 | 2.137 | 6.13 | 1.21 | 5.066 | $P < .01$ |
| WISC | 44 | 104.73 | 11.40 | 1.738 | | | | |
| 9 Binet | 31 | 116.84 | 21.40 | 3.905 | 10.32 | 2.10 | 4.914 | $P < .01$ |
| WISC | 31 | 106.52 | 13.80 | 2.518 | | | | |
| 10 Binet | 29 | 111.66 | 16.55 | 3.129 | 5.69 | 1.74 | 3.270 | $P < .01$ |
| WISC | 29 | 105.97 | 10.85 | 2.051 | | | | |
| 11 Binet | 37 | 108.35 | 15.45 | 2.575 | 6.62 | 1.69 | 3.917 | $P < .01$ |
| WISC | 37 | 101.73 | 10.75 | 1.792 | | | | |
| 12 Binet | 22 | 107.00 | 13.50 | 2.948 | 6.59 | 1.16 | 5.681 | $P < .01$ |
| WISC | 22 | 100.41 | 11.50 | 2.511 | | | | |
| 13 Binet | 30 | 105.33 | 11.65 | 2.161 | 6.66 | 1.30 | 5.123 | $P < .01$ |
| WISC | 30 | 98.67 | 11.55 | 2.143 | | | | |
| 14–15 Binet | 20 | 95.00 | 15.35 | 3.521 | 3.75 | 1.98 | 1.895 | $.10 > P > .05$ |
| WISC | 20 | 91.25 | 14.15 | 3.245 | | | | |
| Total Binet | 332 | 108.45 | 15.85 | 0.871 | 7.22 | 0.50 | 14.440 | $P < .01$ |
| WISC | 332 | 101.23 | 12.75 | 0.701 | | | | |

TABLE 3

MEAN IQ'S AND SIGNIFICANCE OF MEAN DIFFERENCES OF S-B AND WISC VERBAL IQ'S AT EACH AGE LEVEL AND FOR TOTAL GROUP

| Age Level | $N$ | Mean | $SD$ | $SE_M$ | $D$ | $SE_D$ | $t$ | $P$ |
|---|---|---|---|---|---|---|---|---|
| 5 Binet | 38 | 107.26 | 14.90 | 2.451 | 9.34 | 1.90 | 4.916 | $P < .01$ |
| WISC | 38 | 97.92 | 13.20 | 2.170 | | | | |
| 6 Binet | 38 | 111.87 | 14.30 | 2.352 | 7.90 | 1.80 | 4.389 | $P < .01$ |
| WISC | 38 | 103.97 | 15.20 | 2.500 | | | | |
| 7 Binet | 43 | 105.02 | 11.60 | 1.790 | 1.74 | 1.46 | 1.192 | $.30 > P > .20$ |
| WISC | 43 | 103.28 | 10.65 | 1.644 | | | | |
| 8 Binet | 44 | 110.86 | 15.20 | 2.137 | 4.54 | 1.34 | 3.388 | $P < .01$ |
| WISC | 44 | 106.32 | 12.15 | 1.852 | | | | |
| 9 Binet | 31 | 116.84 | 21.40 | 3.905 | 9.84 | 1.99 | 4.945 | $P < .01$ |
| WISC | 31 | 107.00 | 14.70 | 2.682 | | | | |
| 10 Binet | 29 | 111.66 | 16.55 | 3.129 | 3.97 | 1.68 | 2.363 | $.05 > P > .02$ |
| WISC | 29 | 107.69 | 10.40 | 1.966 | | | | |
| 11 Binet | 37 | 108.35 | 15.45 | 2.575 | 3.78 | 1.87 | 2.021 | $.05 > P > .02$ |
| WISC | 37 | 104.57 | 11.95 | 1.975 | | | | |
| 12 Binet | 22 | 107.00 | 13.50 | 2.948 | 2.73 | 2.01 | 1.358 | $.20 > P > .10$ |
| WISC | 22 | 104.27 | 10.50 | 2.293 | | | | |
| 13 Binet | 30 | 105.33 | 11.65 | 2.161 | 2.50 | 1.67 | 1.497 | $.20 > P > .10$ |
| WISC | 30 | 102.83 | 12.20 | 2.263 | | | | |
| 14–15 Binet | 20 | 95.00 | 15.35 | 3.521 | 2.00 | 1.74 | 1.149 | $.30 > P > .20$ |
| WISC | 20 | 93.00 | 16.00 | 3.670 | | | | |
| Total Binet | 332 | 108.45 | 15.85 | .871 | 5.08 | 0.60 | 8.467 | $P < .01$ |
| WISC | 332 | 103.37 | 13.57 | .745 | | | | |

computed between S-B IQ's and those obtained on the WISC Verbal, Performance, and Full Scales at each age level, and for the total group. These correlation coefficients, all of which proved to be significantly different from zero at the .01 level, are presented in Table 1.

In general, the correlations between the S-B Scale and the WISC Full Scale are higher than those obtained between the S-B Scale and either the WISC Verbal or Performance Scales. Of these, the Verbal Scale is more closely correlated with the S-B than is the Performance Scale, except at the youngest age levels.

As a second step in the consideration of the data, the mean IQ's obtained by the pupils at each age level were determined. These data are presented for the S-B Scale and the WISC Full Scale, in Table 2, which also summarizes the results of testing the significance of the obtained mean differences in IQ at each age level.

All but one of the obtained differences between mean IQ's at each age level are significant. As a result, the obtained mean difference of more than 7 IQ points for the total group is also significant. To be sure, a consideration of the results for the WISC Verbal and Performance Scales, presented in Tables 3 and 4, respectively, makes it evident that the obtained differences in mean IQ between the Stanford-Binet and the WISC Full Scale may be attributed, in large part, to such differences as exist between the Stanford-Binet Scale and the WISC Performance Scale.

At almost every age level, the obtained mean IQ's are above 100, indicating that the sample chosen was slightly above average, although it was hoped that an average sample would be obtained by selecting representative schools. It is evident that the small number of

TABLE 4
MEAN IQ'S AND SIGNIFICANCE OF MEAN DIFFERENCES OF S-B AND WISC PERFORMANCE IQ'S AT EACH AGE LEVEL AND FOR TOTAL GROUP

| Age Level | $N$ | Mean | $SD$ | $SE_M$ | $D$ | $SE_D$ | $t$ | $P$ |
|---|---|---|---|---|---|---|---|---|
| 5 Binet | 38 | 107.26 | 14.90 | 2.451 | 12.15 | 1.59 | 7.642 | $P < .01$ |
| WISC | 38 | 95.11 | 12.25 | 2.015 | | | | |
| 6 Binet | 38 | 111.87 | 14.30 | 2.352 | 12.90 | 1.68 | 7.679 | $P < .01$ |
| WISC | 38 | 98.97 | 14.00 | 2.303 | | | | |
| 7 Binet | 43 | 105.02 | 11.60 | 1.790 | 6.62 | 1.85 | 3.578 | $P < .01$ |
| WISC | 43 | 98.40 | 12.05 | 1.860 | | | | |
| 8 Binet | 44 | 110.86 | 15.20 | 2.137 | 9.31 | 2.12 | 4.392 | $P < .01$ |
| WISC | 44 | 101.55 | 12.40 | 1.890 | | | | |
| 9 Binet | 31 | 116.84 | 21.40 | 3.905 | 12.42 | 2.54 | 4.890 | $P < .01$ |
| WISC | 31 | 104.42 | 12.25 | 2.235 | | | | |
| 10 Binet | 29 | 111.66 | 16.55 | 3.129 | 8.63 | 2.72 | 3.173 | $P < .01$ |
| WISC | 29 | 103.03 | 12.50 | 2.363 | | | | |
| 11 Binet | 37 | 108.35 | 15.45 | 2.575 | 9.86 | 2.28 | 4.325 | $P < .01$ |
| WISC | 37 | 98.49 | 12.25 | 2.042 | | | | |
| 12 Binet | 22 | 107.00 | 13.50 | 2.948 | 11.59 | 2.68 | 4.325 | $P < .01$ |
| WISC | 22 | 95.41 | 12.00 | 2.620 | | | | |
| 13 Binet | 30 | 105.33 | 11.65 | 2.161 | 11.83 | 1.70 | 6.959 | $P < .01$ |
| WISC | 30 | 93.50 | 10.80 | 2.004 | | | | |
| 14–15 Binet | 20 | 95.00 | 15.35 | 3.521 | 4.00 | 2.78 | 1.439 | $.20 > P > .10$ |
| WISC | 20 | 91.00 | 13.55 | 3.108 | | | | |
| Total Binet | 332 | 108.45 | 15.85 | 0.871 | 10.17 | 0.72 | 14.250 | $P < .01$ |
| WISC | 332 | 98.28 | 15.05 | 0.827 | | | | |

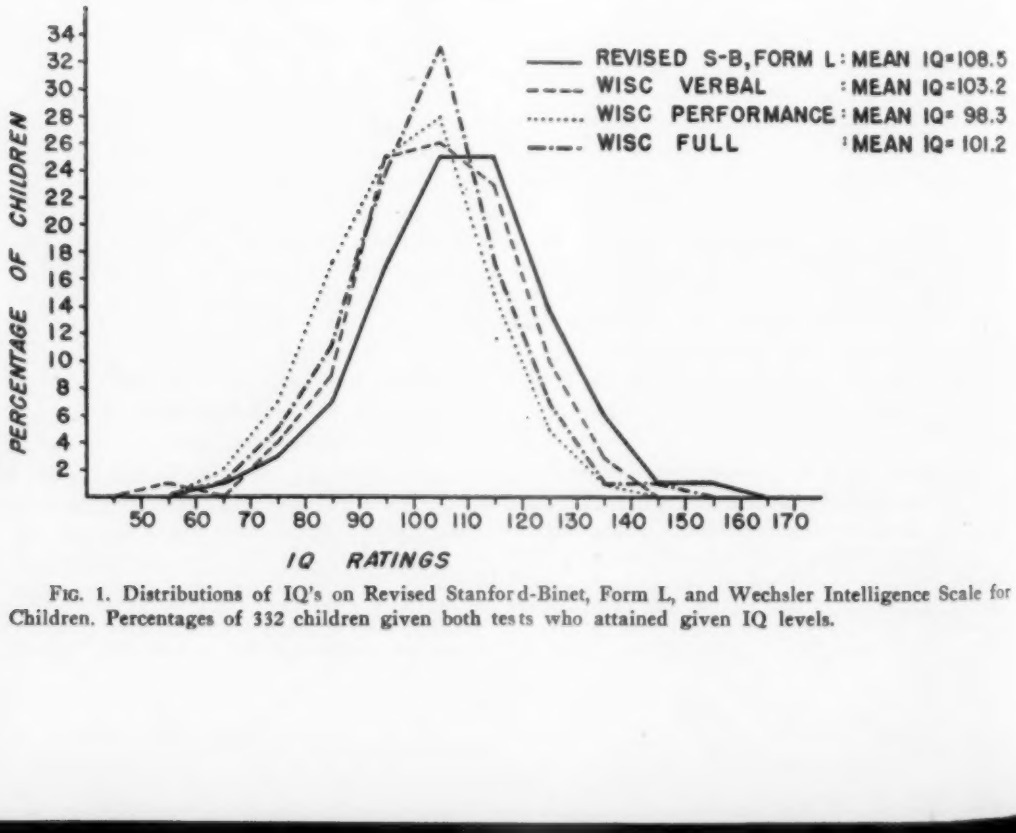


FIG. 1. Distributions of IQ's on Revised Stanford-Binet, Form L, and Wechsler Intelligence Scale for Children. Percentages of 332 children given both tests who attained given IQ levels.

TABLE 5

NUMBER AND PERCENTAGE OF CHILDREN SHOWING GIVEN IQ DIFFERENCES BETWEEN STANFORD-BINET AND WISC VERBAL, PERFORMANCE, AND FULL SCALES

| IQ Difference | WISC Verbal | | WISC Performance | | WISC Full | |
|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | *N* | % |
| Higher on S-B | | | | | | |
| 1–10 points | 123 | 37.0 | 107 | 32.2 | 148 | 44.6 |
| 11–20 points | 74 | 22.3 | 77 | 23.2 | 85 | 25.6 |
| 21 or more points | 15 | 4.5 | 68 | 20.5 | 20 | 6.0 |
| Total | 212 | 63.9 | 252 | 75.9 | 253 | 76.2 |
| Higher on WISC | | | | | | |
| 1–10 points | 90 | 27.1 | 58 | 17.5 | 60 | 18.1 |
| 11–20 points | 16 | 4.8 | 14 | 4.2 | 6 | 1.8 |
| Total | 106 | 31.9 | 72 | 21.7 | 66 | 19.9 |
| No difference | 14 | 4.2 | 8 | 2.4 | 13 | 3.9 |

cases utilized at each age level made it virtually impossible to secure complete representativeness. It should be noted, too, that the selection of schools was based upon group-test results which, as many comparative studies have indicated, do tend to give results several points lower than individual examinations. In this connection, too, it may be well to observe that Terman's standardization groups generally showed mean IQ's above 100 [2].

Table 3 presents a similar comparison between the obtained mean IQ's for the S-B and WISC Verbal Scales. For the total group, a difference of approximately 5 points in IQ was obtained between the S-B and the WISC Verbal Scales. This difference proved to be significant at the .01 level. Examination of the obtained results at each age level indicates that, for the most part, smaller differences appear. In large measure, the results for the total group may be attributed to the relatively large differences noted between the test results at the lower age levels.

Table 4 presents a similar comparison between the obtained mean IQ's for the S-B and WISC Performance Scales. Not only is the mean difference in IQ significant for the total group, but the same situation holds true in all but one of the instances in which separate age levels were considered.

Consideration of the results presented in Tables 2, 3, and 4 indicates that the means for the total group are all significantly higher on the Revised S-B, Form L, than on any of the WISC Scales. At every age level, the WISC Verbal Scale mean IQ is closer to the S-B mean IQ than either the WISC Full Scale or Performance Scale mean IQ. At every age level, too, the WISC Performance Scale mean IQ is lower than the WISC Verbal Scale IQ.

Distributions of the IQ's on each of the WISC Scales and on the S-B are presented in Figure 1. Inspection of the data upon which the distributions presented in Figure 1 are based reveals that the range of results on the S-B is greater than that on the WISC Scales at the upper levels. The highest IQ's observed on the WISC were 145 on the Verbal Scale, 143 on the Full Scale, and 136 on the Performance Scale, compared with 167 on the S-B. Throughout the distribution, the WISC Performance Scale shows the greatest discrepancy from the S-B.

The distributions presented in Figure 1 are concerned with group results. A summary of individual differences in IQ's obtained on the two examinations is presented in Table 5. When one compares the S-B results with any of the WISC Scale results, one finds that the large majority of children obtained higher results on the S-B. Thus, on the S-B Scale, approximately 76 per cent of the group of 332 children were rated higher than on the WISC Full Scale or Performance Scale and 65 per cent were rated higher than on the WISC Verbal Scale.

Conversely, a minority of the group obtained higher ratings on the WISC Scales — approximately 20 per cent on the Full Scale and Performance Scale, and approximately 32 per cent on the Verbal Scale.

Analysis of the size of the differences reveals wide variations in individual cases. Thus, for the WISC Full Scale, differences ranged from 12 points higher on the WISC to 38 points higher on the S-B. For the Performance Scale, differences were even wider, ranging from 20 points higher on the WISC to 54 points higher on the S-B.

On the WISC Full Scale and Verbal Scale, approximately two-thirds of the group obtained IQ's which were within 10 points of their S-B IQ's, whereas on the Performance Scale, on which there were greater differences, slightly more than one-half of the group obtained IQ's within 10 points of their S-B IQ's. One-fifth of the group obtained S-B IQ's 21 or more points higher than their WISC Performance IQ's, as compared with approximately one-twentieth of the group on the other WISC Scales.

Although the preceding analysis has indicated that one may expect differences to appear between WISC and Stanford-Binet IQ's, and that the latter would tend to be the higher score for the large majority of children, an additional question may be posed: Will the greater differences between WISC IQ's and Stanford-Binet IQ's tend to be associated with higher or lower Stanford-Binet scores? In order to determine the answer to this question, correlations were computed between Stanford-Binet IQ and obtained differences between Stanford-Binet IQ and WISC Verbal, WISC Performance, and WISC Full Scale IQ's. These data are presented in Table 6.

TABLE 6

CORRELATION BETWEEN STANFORD-BINET IQ AND DIFFERENCES BETWEEN STANFORD-BINET IQ AND WISC VERBAL WISC PERFORMANCE, AND WISC FULL SCALE IQ'S

| Age Level | N | S-B and Diff. between S-B and WISC Verbal | S-B and Diff. between S-B and WISC Performance | S-B and Diff. between S-B and WISC Full Scale |
|---|---|---|---|---|
| 5 | 38 | .363* | .541** | .449** |
| 6 | 38 | .397* | .387* | .396* |
| 7 | 43 | .429** | .365* | .545** |
| 8 | 44 | .604** | .564** | .636** |
| 9 | 31 | .553** | .773** | .691** |
| 10 | 29 | .683** | .630** | .683** |
| 11 | 37 | .434** | .672** | .659** |
| 12 | 22 | .468* | .448* | .415 |
| 13 | 30 | .314 | .597** | .438* |
| 14–15 | 20 | .139 | .449* | .239 |
| Total | 332 | .546** | .625** | .533** |

*Significantly different from zero at the .05 level.
**Significantly different from zero at the .01 level.

There is a definite tendency for greater differences between Stanford-Binet and WISC IQ's to be associated with the higher Stanford-Binet IQ's. Not only may this generalization be advanced for the total group of 332 pupils, but it holds true at almost every age level. In relatively few instances do the obtained correlation coefficients fail to reach an acceptable level of confidence.

An analysis of the data for the WISC Full Scale for the total group of 332 children is presented in Table 7, which gives the percent-

TABLE 7

PERCENTAGE OF CHILDREN AT GIVEN STANFORD-BINET IQ LEVELS SHOWING GIVEN DIFFERENCES BETWEEN STANFORD-BINET AND WISC FULL SCALE IQ'S

| | | Higher on S–B | | | Higher on WISC | | | |
|---|---|---|---|---|---|---|---|---|
| S-B IQ | N | 1–10 points | 11–20 points | 21 or more points | 1–10 points | 11–20 points | No Diff. | Mean Diff. (S-B — WISC) |
| 130 and over | 28 | 14.3 | 46.4 | 35.7 | | | 3.6 | +19.3 |
| 120–129 | 46 | 32.6 | 45.7 | 10.9 | 4.3 | | 6.5 | +10.5 |
| 110–119 | 83 | 59.0 | 25.3 | 6.0 | 9.6 | | 0.0 | + 8.5 |
| 100–109 | 84 | 45.2 | 23.8 | | 25.0 | 2.4 | 3.6 | + 4.7 |
| 90– 99 | 56 | 44.6 | 16.1 | | 30.4 | 1.8 | 7.1 | + 2.6 |
| 80– 89 | 23 | 47.8 | 4.3 | | 34.8 | 8.7 | 4.3 | + 0.7 |
| Below 80 | 12 | 50.0 | | | 33.3 | 8.3 | 8.3 | + 0.2 |

age of children at each IQ level who were higher on the S-B or WISC, as well as the number of points higher.

Although the number of cases with IQ's below 90 and above 130 are small, the trend throughout is consistent in indicating that differences tend to increase with increase in IQ level. Thus, the mean differences increase from 0.2 points higher on the S-B for those who had S-B IQ's below 80, to 4.7 points higher on the S-B for those with S-B IQ's of 100 to 109, to 19.3 points higher on the S-B for those who had S-B IQ's of 130 or higher. Only children with S-B IQ's of 110 or higher showed differences, in favor of the S-B, of 21 points or more.

Throughout, there were fewer children who showed higher ratings on the WISC Full Scale than on the S-B. When such children are noted, they tend to be found more frequently at the lower S-B IQ levels, and the difference between the S-B and WISC IQ's is generally small. The lower the IQ, then, the more likely the WISC and S-B ratings agree.

An additional factor is worthy of consideration, and may also be expressed in the form of a question: Is there any relationship between chronological age and difference between IQ's obtained on the two tests? The answer, in this instance, may be determined through the use of the chi square technique. A cross-tabulation of chronological ages and IQ differences between the Stanford-Binet and WISC Verbal Scales is presented in Table 8.

TABLE 8

CROSS-TABULATION OF CHRONOLOGICAL AGE AND DIFFERENCE BETWEEN STANFORD-BINET AND WISC VERBAL SCALE IQ'S

| | IQ Differences (S-B – WISC Verbal) | | | | | | |
|---|---|---|---|---|---|---|---|
| Age Level | −6 and over | −1 to −5 | 0 to +4 | +5 to +9 | +10 to +14 | +15 and over | Total |
| 5–6 | 8 | 5 | 13 | 16 | 10 | 24 | 76 |
| 7–8 | 16 | 16 | 13 | 19 | 15 | 8 | 87 |
| 9–10 | 12 | 6 | 9 | 10 | 11 | 12 | 60 |
| 11–12 | 9 | 15 | 11 | 11 | 6 | 7 | 59 |
| 13–15 | 12 | 7 | 11 | 10 | 7 | 3 | 50 |
| Total | 57 | 49 | 57 | 66 | 49 | 54 | 332 |

$\chi^2 = 35.228$
$.02 > P > .01$

A significant relationship exists between CA and difference between Stanford-Binet and WISC Verbal IQ's. In general, the younger the child, the greater the difference in favor of the Stanford-Binet. The results of this approach, to be sure, confirm that growing out of the data summarized in Table 3.

TABLE 9

CROSS-TABULATION OF CHRONOLOGICAL AGE AND DIFFERENCE BETWEEN STANFORD-BINET AND WISC PERFORMANCE SCALE IQ'S

| | IQ Differences (S-B–WISC Performance | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Age Level | −6 and over | −1 to −5 | 0 to +4 | +5 to +9 | +10 to +14 | +15 to +19 | +20 to +24 | +25 and over | Total |
| 5–6 | 2 | 7 | 3 | 18 | 12 | 16 | 9 | 9 | 76 |
| 7–8 | 13 | 10 | 16 | 17 | 7 | 9 | 5 | 10 | 87 |
| 9–10 | 8 | 5 | 8 | 8 | 10 | 6 | 4 | 11 | 60 |
| 11–12 | 5 | 10 | 8 | 10 | 7 | 3 | 8 | 8 | 59 |
| 13–15 | 7 | 5 | 6 | 12 | 5 | 4 | 6 | 5 | 50 |
| Total | 35 | 37 | 41 | 65 | 41 | 38 | 32 | 43 | 332 |

$\chi^2 = 36.943$
$.20 > P > .10$

Table 9 presents a similar cross-tabulation of CA's and IQ differences between the Stanford-Binet and WISC Performance Scales. Analysis of the data fails to reveal a significant difference between CA and IQ differences. Here, too, the obtained results parallel those reported in Table 4.

TABLE 10

CROSS-TABULATION OF CHRONOLOGICAL AGE AND DIFFERENCE BETWEEN STANFORD-BINET AND WISC FULL SCALE IQ'S

| | IQ Differences (S-B – WISC Full) | | | | | |
|---|---|---|---|---|---|---|
| Age Level | −1 and over | 0 to +4 | +5 to +9 | +10 to +14 | +15 and over | Total |
| 5–6 | 4 | 10 | 17 | 21 | 24 | 76 |
| 7–8 | 24 | 17 | 24 | 8 | 14 | 87 |
| 9–10 | 11 | 11 | 16 | 8 | 14 | 60 |
| 11–12 | 19 | 8 | 14 | 9 | 9 | 59 |
| 13–15 | 9 | 13 | 12 | 9 | 7 | 50 |
| Total | 67 | 59 | 83 | 55 | 68 | 332 |

$\chi^2 = 35.731$
$P < .01$

Table 10 presents a third cross-tabulation —that between CA and IQ differences between the Stanford-Binet and WISC Full Scale IQ's. In this instance, as well as that reported for the WISC Verbal Scale, the obtained relationship is a statistically significant one. Here, too, the indications are that younger children will tend to show a larger difference, in favor of Stanford-Binet IQ, between Stanford-Binet and WISC Full Scale IQ's than older children.

## CONCLUSIONS

1. For the total group of 332 pupils tested, correlations of .739, .644, and .817 were obtained between the Stanford-Binet, Form L, and the WISC Verbal, Performance, and Full Scales, respectively.

2. Correlations between Stanford-Binet and WISC Full Scale IQ's vary between approximately .75 and .90 at various age levels. These tend to be somewhat higher than those obtained between the Stanford-Binet and the WISC Verbal Scale IQ's, which vary from approximately .65 to .90. The latter, in turn, are higher than those noted between Stanford-Binet and WISC Performance Scale IQ's, which range from approximately .50 to .75.

3. The mean IQ's on the Revised Stanford-Binet, Form L, are higher than those on the Wechsler Intelligence Scale for Children. The widest discrepancy is between the Stanford-Binet and the WISC Performance Scale. The WISC Verbal Scale IQ's are closer to the Stanford-Binet (mean difference 5.1 points) than either the WISC Performance (mean difference 10.2 points) or Full Scale IQ's (mean difference 7.3 points). For the 332 children between 5½ and 15½ years in this study, the mean IQ's were 108.5 on the Stanford-Binet, 103.4 on the WISC Verbal Scale, 98.3 on the WISC Performance Scale, and 101.2 on the WISC Full Scale.

4. These obtained differences between Stanford-Binet and WISC Full Scale IQ's proved to be significant at all but one age level. Differences between Stanford-Binet and WISC Verbal Scale IQ's tended to be significant only at the younger age levels. Differences between Stanford-Binet and WISC Performance Scale IQ's are significant at all but one age level. It is evident that the obtained differences in mean IQ between the Stanford-Binet and WISC Full Scale may be attributed, in large part, to such differences as exist between the Stanford-Binet Scale and the WISC Performance Scale.

5. The study of individual results showed that approximately two-thirds of the children obtained Stanford-Binet ratings within 10 points of those on the WISC Full Scale or Verbal Scale, as against approximately one-half on the WISC Performance Scale.

6. There is a definite tendency for greater differences between Stanford-Binet and WISC IQ's to be associated with the higher Stanford-Binet IQ's. This generalization may be advanced at almost every age level, and for differences between the Stanford-Binet and WISC Verbal, Performance, and Full Scales.

7. The indications are that differences between Stanford-Binet IQ and WISC Verbal and Full Scale IQ's tend to be associated with chronological age, in that such differences are larger at younger age levels. The same statement cannot be made for IQ differences between the Stanford-Binet and WISC Performance Scales.

## DISCUSSION

In the present study, the WISC gave results similar to the Revised Stanford-Binet, Form L, in the large majority of cases at the lower IQ levels, but the discrepancies appearing at the upper IQ levels may be considered too large to permit the use of the WISC in place of the S-B until further work has been done. It is true that larger discrepancies are to be expected at higher IQ levels than lower ones — Terman and Merrill report average differences between the two forms of the Revised Stanford-Binet as 5.92 for IQ's above 130, as compared with 2.49 for IQ's below 70. In the present study, however, the discrepancies between the WISC and S-B are much larger (19.3 for IQ's above 130; 10.5 for IQ's between 120 and 129). Since the number of cases involved is small, further comparisons are necessary to see if these findings will be confirmed.

The question can be raised as to whether the discrepancies are due to underestimation of ability by the WISC or overestimation by the S-B, or to a combination of such under- and overestimation. Only widespread work with

the WISC, comparisons of distributions of large numbers of representative groups, and clinical findings can provide an answer to this question.

During the period before such data are available and before the evidence is conclusive, psychologists will and should certainly use the WISC, but will still feel the need for checking with the Stanford-Binet in those instances where clinical judgment and WISC results seem at variance.

There is no doubt that the Wechsler Intelligence Scale for Children is a valuable addition to the psychologist's clinical techniques. The inclusion of performance material as an integral part of the Full Scale, the possibility of obtaining both a verbal and performance rating, and the point scale approach, open possibilities for clinical uses that should be as fruitful at the children's level as they have proved to be on the adult scales.

*Received March 5, 1951.*

REFERENCES

1. SEASHORE, H., WESMAN, A., AND DOPPELT, J. The standardization of the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1950, **14**, 99-110.
2. TERMAN L. M., AND MERRILL, MAUD A. ***Measuring intelligence.*** Boston: Houghton Mifflin, 1937.

# A COMPARISON OF MASCULINITY-FEMININITY MEASURES

BERNARD F. SHEPLER
FLORIDA STATE UNIVERSITY

THE Franck Completion Test [2], a projective method for measuring masculinity-femininity (hence, called MF), was recently introduced to the clinical field. Because the initial results looked promising, additional experimentation is indicated. Franck [1] suggested that one of the areas needing research was comparison with other measures of MF. The purpose of this experiment is to compare the Franck Test with three standard measures of MF.

The experimental group was composed of 57 male and 67 female students, a 100 per cent sample of three classes in general psychology. Owing to requirements of the University, 95 per cent of all undergraduate students enroll in general psychology. For this reason it is felt that the experimental group may be considered an adequately random sample.

Each subject was administered a battery of four measures of MF. The four components of this battery were: (1) the MF Scale of the Strong Vocational Interest Blank [5], Form M for men [6] and Form W for women [7]; (2) the *Mf* Scale of the MMPI [3, 4]; (3) the Terman-Miles Attitude-Interest Analysis Test [8, 9]; and (4) the Franck Completion Test.

Form M and Form W of the Strong each consist of 400 items. Items are checked "Like," "Indifferent," "Dislike," or, at times, are ranked in order of preference. The *Mf* Scale of the MMPI is made up of 60 questions from a total group of 550. Questions are answered "True," "False," or "Cannot Say." The Terman-Miles Test (Form A) contains 456 items divided into seven subtests. It is a pencil-and-paper test of the questionnaire variety, and items are checked in a similar manner to those on the Strong. The Franck Test is made up of 36 incomplete drawings, each containing a few simple lines. Each stimulus extends over approximately one square inch, located in the center of a cell two and one-quarter inches square. Subjects are instructed to complete the drawings in any manner that is enjoyable.

The *F*-test was applied to determine whether the experimental sample was similar to the normative groups used by the authors of the various measures. To determine the reliability of scoring the Franck, Pearson product-moment correlations were computed for both male and female groups as rated by three scorers.

Critical ratios were calculated for the four measures of MF to determine the ability of these measures to discriminate between males and females per se.

Correlations were computed between the four measures of MF for both male and female groups; correlations between the Franck and the three standard instruments were computed to determine to what extent the four instruments were measuring the same aspects of MF; correlations between the three standard instruments were computed for additional comparative puposes.

TABLE 1

MEANS AND SD'S FOR MALES AND FEMALES OF EXPERIMENTAL GROUPS AND TEST AUTHOR'S GROUPS

| MF Scale | Group | Men's N | Men's Mean | Men's SD | Women's N | Women's Mean | Women's SD |
|---|---|---|---|---|---|---|---|
| Terman-Miles | Exp. | 57 | 62.17 | 55.50 | 67 | –57.11 | 45.45 |
| | Author | 130 | 67.42 | 47.65 | 130 | –60.81 | 39.15 |
| Strong | Exp. | 57 | 58.15 | 79.80 | 67 | –39.53 | 26.8 |
| | Author | 143 | 60.6 | 81.72 | 150 | –43.4 | 23.0 |
| MMPI | Exp. | 57 | 24.92 | 5.62 | 67 | 34.97 | 5.30 |
| | Author | 117 | 20.44 | 5.13 | 108 | 36.51 | 4.83 |
| Franck | Exp. | 57 | 16.82 | 2.96 | 67 | 19.79 | 3.38 |
| | Author | 150 | 14.41 | 3.39 | 150 | 19.61 | 3.28 |

## RESULTS

Means and standard deviations for males and females, experimental groups and test authors' groups, are presented in Table 1. Results of the *F*-test are not included; all failed to meet requirements for rejection of the null hypothesis at the 5 per cent level of confidence.

The correlations between scorers for both male and female groups are not included, but all were above .90. All were significant at the 1 per cent level of confidence indicating satisfactory reliability of scoring.

TABLE 2

*CR*'s BETWEEN MALES AND FEMALES*

| MF Scale | *CR* |
|---|---|
| Terman-Miles | 12.35 |
| Strong | 8.68 |
| MMPI | 10.10 |
| Franck | 5.19 |

*Experimental groups only.

Table 2 presents the critical ratios that were computed to determine the ability of each of the four measures to discriminate between males and females per se. All four mean differences are significant at the 1 per cent level.

Correlations between the four measures of MF both males and females are given in Table 3. The negative correlations found in Table 3 are due to scoring peculiarities. A "high" score is masculine on the Strong and Terman-Miles, whereas a "low" score is masculine on the MMPI and Franck. The opposite obtains for feminine scores.

TABLE 3

CORRELATIONS BETWEEN MF MEASURES FOR MALES AND FEMALES

| | Franck vs. T-M | Franck vs. Strong | Franck vs. MMPI | Strong vs. T-M | Strong vs. MMPI | MMPI vs. T-M |
|---|---|---|---|---|---|---|
| Males | −.087 | .107 | −.001 | .565 | −.499 | −.656 |
| Females | −.013 | −.025 | .149 | .669 | −.551 | −.534 |

As indicated in Table 3, correlations between the Franck and the standard measures were not significant. The correlations between the standard measures were all significant from zero at the 1 per cent level.

## INTERPRETATIONS

Results of the *F*-test indicate that the experimental sample is similar to the normative samples, indicating equivalent randomness has been obtained.

The high correlation between scorers on the Franck may be accounted for by the manner in which the tests were scored. Whenever a discrepancy occurred between scorers, a conference was held and agreement reached if possible. In many instances agreement could not be reached, with the result that as many as ten item differences between scorers appeared on a single test. When the total score of the individual was computed, these differences canceled each other out in most instances. In no case was the discrepancy of the totals greater than 2.

Despite the high reliability of scoring, it is felt that the scoring manual is in need of refinement. While some items appeared to be relatively simple to score, others were very difficult due to lack of clarity of directions. Examples of "typical" male and female completions were confusing owing to similarity. The seemingly obvious lack of discriminatory ability of some of the stimulus drawings indicates that an item analysis might prove to be fruitful.

The data presented in Tables 2 and 3 are thought-provoking. From Table 2 we see that the Franck does as adequate a job of discriminating between male and female groups as do the other measures, i.e., critical ratios for all four measures are significant at the 1 per cent level. Table 3 suggests that the Terman-Miles, the Strong, and the MMPI are measuring many aspects of MF that are common to all three instruments. All correlations between these three instruments are significant at the 1 per cent level.

In contradistinction to this, the Franck does *not* correlate with any of the three standard measures. All correlations are found not to be significant. Taking these facts into consideration, it would seem that the Franck is measuring entirely different aspects of MF from those measured by the other three instruments. What these aspects of MF are is unknown and beyond the scope of this paper. Further research is indicated in the light of these findings.

## CONCLUSIONS

1. Critical ratios between a group of males and a group of females indicates that the Franck Completion Test has the ability to discriminate between males and females per se. This discriminatory ability is significant at the 1 per cent level of confidence. Therefore, it may be assumed that males and females complete the stimulus drawings in a different manner.

2. Nonsignificant correlations between the Franck and the Terman-Miles, the *Mf* Scale of the MMPI, and the MF scale of the Strong indicate that the Franck is apparently measuring different aspects of MF from those measured by these three instruments.

3. Further research is needed to determine what these aspects of MF are that the Franck appears to be measuring.

4. Indications are that the scoring manual of the Franck is in need of refinement if the test is to be used to its fullest advantage and promise.

5. The need for an item analysis of the Franck is suggested.

*Received March 16, 1951.*

## REFERENCES

1. FRANCK, KATE. *Manual for Completion Test (Masculinity-femininity Scale).* Berkeley, California: E. S. Mimopolus, 2254 Telegraph Avenue, 1949. (Mimeo.)
2. FRANCK, KATE, AND ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, **13**, 247-256.
3. HATHAWAY, S. R., AND MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
4. HATHAWAY, S. R., AND MCKINLEY, J. C. *Supplementary manual for the Minnesota Multiphasic Personality Inventory (for group form).* New York: Psychological Corp., 1946.
5. STRONG, E. K., JR. *Vocational interests of men and women.* Stanford, Calif.: Stanford Univ. Press, 1945.
6. STRONG, E. K., JR. *Manual for Vocational Interest Blank for Women.* Stanford, Calif.: Stanford Univ. Press, 1947.
7. TERMAN, L. M., AND MILES, CATHARINE C. *Sex and personality.* New York: McGraw-Hill, 1936.
8. TERMAN, L. M., AND MILES, CATHARINE C. *Manual of information and directions for use of the Attitude-Interest Analysis Test.* New York: McGraw-Hill, 1938.

# THE SEMANTICS OF INTELLECTUAL CLASSIFICATION

RAYMOND CORSINI
WISCONSIN DEPARTMENT OF PUBLIC WELFARE

A COMMON problem of scientists is the conversion of quantitative symbols into qualitative ones in some manner, to permit those individuals who may be uninformed about the technical meanings of specialized numbers to understand their import through the use of words. In psychological practice, this problem is a real and continuing one, and it is this writer's intention to propose a specific, rational solution to one case.

Intelligence test results commonly are converted into IQ's and then into statements such as "average," "dull," and "superior." There are several reasons why use of such classificatory terms is preferable to the use of IQ's, centiles, standard scores, etc. One reason is that the use of classification terms implies a general "band" of values rather than a specific point-value, and, consequently, does not imply the precision that a number might. A further reason is that a more accurate impression as well as a more comprehensive one may be communicated to statistically naive individuals when general terms are used. As we shall show, however, some very real possibilities of misinterpretation can occur, perhaps in unexpected directions, by our usual systems of intellectual classification. We have attempted to present some data which in method and content may have value in resolving this problem, remembering that the efforts of test constructors and of clinical psychologists come to naught if the ultimate user of psychometric data should misapprehend them through semantic difficulties.

Let us consider two standard schemes for the transmutation of IQ's into classificatory terms, the well-known systems of L. M. Terman [1] and David Wechsler [2]. Exact correspondence between these two schemes exists in two respects: both systems have seven "bands" of values, and the middle portion in both cases is called "average" and contains 50 per cent of the entire distribution. No other agreement exists. But some of the differences need not be considered trivial. "Superior" means IQ 110 to 120 for Terman, and IQ 120 to 127 for Wechsler. The lowest class has an upper limit of an IQ of 70 for Terman, and 65 for Wechsler;[1] the highest class starts with an IQ of 140 for Terman, and 128 for Wechsler, etc. [1, p. 79: 2, p. 37].

Both these schemes are admitted by their originators to be tentative, Terman referring to his system as "suggested" [2, p. 79] and Wechsler stating in reference to his system, "some agreement both as to system of notation, and interpretation of results, is obviously necessary. Such agreement could best be achieved through conventions established by an international meeting of psychologists and psychiatrists" [2, p. 39]. It is quite obvious that both schemes were established by fiat and that each represents but one man's opinion. There are several problems that appear: (1) How many "bands" of value are optimal? (2) What should the limits for these "bands" be? (3) How shall they be labeled? Undoubtedly, each of these questions can be settled by "international conventions," but a more direct and satisfactory solution might be to attempt an empirical investigation using the familiar techniques of psychometrics. In this paper we shall discuss a rational solution to the three questions.

## THE METHOD

The 10 terms suggested by Terman and Wechsler, as well as 49 others obtained from diverse sources, were reproduced on a sheet of legal-size paper. Beneath each term a 100-millimeter line was placed. Instruction sheets

[1]Wechsler quotes IQ 65 as the upper limit of "defective" on page 40 and page 190, but sets it at 67 on page 227 [2].

were prepared which read as follows:

This is an experiment to determine what certain words, descriptive of intelligence, mean to you. You are to rate each word as you understand it. Notice that below each word there is a horizontal line. You are to make a mark on the line under every word to convey the impression the word makes on you. Consider the line to represent all people. Making a mark at the exact center would mean "average" or "normal" intelligence, the intelligence possessed by the person exactly in the middle. Making a mark at a point halfway between the center and the right hand end represents intelligence at the 75 per cent point. Making a mark at a point one-tenth of the length of the entire line from the lower end means, of course, intelligence as represented by people at the 10 per cent point.

To summarize, make your mark where you feel it belongs in describing people's intelligence. View the line as a percentage indicator with 0 per cent and 100 per cent at the ends, 50 per cent in the middle, and 25 per cent and 75 per cent halfway between the left end and the center and halfway between the center and the right end.

## SUBJECTS

Five populations of judges of ten individuals each were used to attempt to get sufficient judges of each type, and a rather wide diversity. Group A consisted of intelligent adults, acquaintances of the writer; Group B consisted of inmates of a New York State Prison who had been assigned to the prison school as teachers of elementary and high school courses; Group C consisted of psychologists, four being clinical psychologists and six being graduate students at Syracuse University; Group D was made up of sociologists — two parole officers, two probation officers, and six social workers; Group E consisted of high school teachers employed in the Auburn, New York, School System.[2]

Each subject's rating of the 59 terms was measured by a 100-millimeter ruler, and was posted. Means and mean deviations were computed for each term for each of the five groups and for the total group of 50 subjects. Each term was ranked, and rank-order correlations were made. Because the correlation in each of the ten pairs exceeded a $\rho$ of .99 and because by inspection no gross constant differences in means occurred between groups, the 50 judges were considered to form as homogeneous a population as any of the separate groups, and were, therefore, treated as one group.

[2]The writer wishes to thank the following people who assisted him by obtaining judges: Dr. Herbert Sanderson, of the Jewish Community Service Society; Miss Regine Seidler, of the Des Moines Child Guidance Service; Mr. Larry Argiro, then of the Auburn, New York, School System; Mr. Bernard Kogon, then of the New York School of Social Welfare; and Mr. David Guralnick, of the New York State Employment Service.

## RESULTS

The 59 terms contained some that were unique and awkward, such as "at least average intelligence," "limited intelligence," "low order of intelligence." Some of the terms were inversions and variations on the same theme, such as "retarded mental development," "mentally retarded," "retarded intelligence." Eliminating these two classes left 23 terms, including the ten suggested by Terman and Wechsler. Table 1 lists these 23 terms together with their mean values and their mean deviations for our 50 judges — the terms being arranged in rank order. We note some unexpected facts. Terman placed "dullness" at IQ of 80 to 90 and "borderline deficiency" at IQ of 70 to

TABLE 1

MEANS AND MEAN DEVIATIONS OF RATINGS OF INTELLECTUAL CLASSIFICATIONS

| Classification | Mean Value | Mean Deviation |
|---|---|---|
| Idiot | 2.5 | 2.6 |
| Imbecile | 5.8 | 4.3 |
| Feebleminded | 11.0 | 6.4 |
| Defective | 16.0 | 8.3 |
| Moron | 17.0 | 8.0 |
| Inferior | 28.0 | 8.1 |
| Backward | 28.0 | 7.6 |
| Dull | 31.0 | 8.3 |
| Limited | 33.0 | 7.9 |
| Borderline | 34.0 | 10.3 |
| Retarded | 34.0 | 7.5 |
| Low Average | 37.0 | 5.3 |
| Below Average | 37.0 | 5.5 |
| Dull Normal | 37.0 | 6.7 |
| Average | 50.0 | 1.7 |
| Normal | 50.0 | 3.2 |
| Bright Normal | 59.0 | 5.7 |
| Bright | 65.0 | 8.4 |
| High Average | 67.0 | 7.7 |
| Superior | 76.0 | 8.8 |
| Very Bright | 77.0 | 6.4 |
| Very Superior | 90.0 | 5.2 |
| Genius | 96.0 | 3.6 |

TABLE 2
MEAN RATINGS AND MEAN DEVIATIONS OF CLASSIFICATION TERMS BY FIVE POPULATIONS OF TEN JUDGES EACH

(A, Intelligent Adults; B, Prison Inmate Teachers; C, Psychologists; D, Sociologists; E, High School Teachers.)

| Classification | A | | B | | C | | D | | E | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *MD* | *M* | *MD* | *M* | *MD* | *M* | *MD* | *M* | *MD* |
| Genius | 95 | 5.0 | 98 | 1.8 | 94 | 3.8 | 94 | 4.5 | 98 | 3.0 |
| Very Superior | 91 | 5.6 | 89 | 8.4 | 91 | 4.2 | 85 | 5.6 | 96 | 2.0 |
| Superior | 77 | 9.1 | 72 | 11.6 | 80 | 5.4 | 70 | 8.6 | 80 | 9.5 |
| High Average | 69 | 11.5 | 68 | 8.3 | 67 | 3.2 | 66 | 7.6 | 69 | 7.9 |
| Bright Normal | 51 | 4.0 | 61 | 7.8 | 62 | 3.9 | 57 | 3.3 | 65 | 9.6 |
| Average | 50 | 0.5 | 48 | 4.1 | 50 | 0.0 | 50 | 1.1 | 51 | 2.6 |
| Dull Normal | 38 | 7.4 | 34 | 6.6 | 32 | 7.7 | 39 | 7.8 | 42 | 4.1 |
| Low Average | 33 | 6.5 | 31 | 7.5 | 42 | 2.0 | 43 | 3.5 | 35 | 7.0 |
| Borderline | 38 | 11.8 | 33 | 15.8 | 28 | 7.8 | 39 | 7.4 | 41 | 8.5 |
| Dull | 29 | 9.0 | 27 | 8.3 | 32 | 7.0 | 32 | 8.8 | 37 | 8.2 |
| Inferior | 31 | 9.4 | 27 | 8.8 | 28 | 5.0 | 28 | 8.0 | 25 | 9.5 |
| Defective | 21 | 12.6 | 10 | 6.2 | 18 | 9.4 | 17 | 6.7 | 16 | 6.7 |
| Feebleminded | 7.6 | 6.8 | 6.1 | 4.9 | 14 | 6.8 | 17 | 7.7 | 12 | 5.8 |
| Imbecile | 2.3 | 2.4 | 2.0 | 4.0 | 9.8 | 3.8 | 8.9 | 5.7 | 6.1 | 5.9 |

80. Wechsler rated "dull normal" at 80 to 90 and "borderline" at 66 to 79. Kuhlmann placed "dull" at 85 to 94 and "borderline" at 75 to 84. We find, however, that "borderline" has a higher mean value, as far as our judges go, than "dull," and that "dull" and "dull-normal" have quite different connotations. "Bright normal" and "dull normal," which were placed symmetrically above and below the mean by Wechsler, are respectively 9 points and 13 points away from the "average." Terms near the extremes are more homogeneous than terms near the center. If we take the *MD* of "average" to indicate mechanical error, we find that "idiot" has 1.5 times as much variation, "genius" 2.1, while "moron" has 4.7, "superior" 5.1, and "borderline," the most heterogeneous term, has a *MD* 6.1 as large as "average."

Table 2 contains the mean values and mean deviations of the ten terms suggested by Terman and Wechsler plus four others for our five groups of judges. We note the psychologists' averages exceed the average for the total group in eight cases, are exactly equal in two cases, and are lower in four cases. By inspection, we notice some differences, but they do not appear to be in any direction to make the formidable task of analysis of significance worthwhile. It is of interest to note that "borderline" has, most probably, a significantly lower value for the psychologists than for other classes, and this is also the case for "dull normal," which undoubtedly reflects the influence of special training, and which implies, insofar as these words are concerned in this context, that psychologists are using familiar words in a special way.

### A NEW SCALE

Can we compare our rating values with any known scale, so that we can give these terms empiric IQ values? This is a moot point, for although the directions for the judges implied a centile scale, whether or not we can so consider it, is another question. Perhaps we could for the psychologists, but inasmuch as our aim is not to find what one special group thinks, but what a diverse sample of people feels, we see no point in using the psychologists alone. Accepting, tentatively, the hypothesis that our rating scale is a centile scale, and that our group mean values are mean affective centile values, when we arrange these terms on a centile scale and compare their placement as suggested by Terman and Wechsler, we find in Figure 1 that the values by fiat and the values by experiment differ considerably in some cases. The area from 25 per cent to 75 per cent has been reserved by both Terman and

Wechsler for the range of "average," but our judges place "dull," "borderline," "dull normal," and "bright normal" within this range. If we look back at Table 1, we find a total of 14 terms from "inferior" to "high average" being included in the established range of "average," and even "superior" comes to within one point of inclusion!

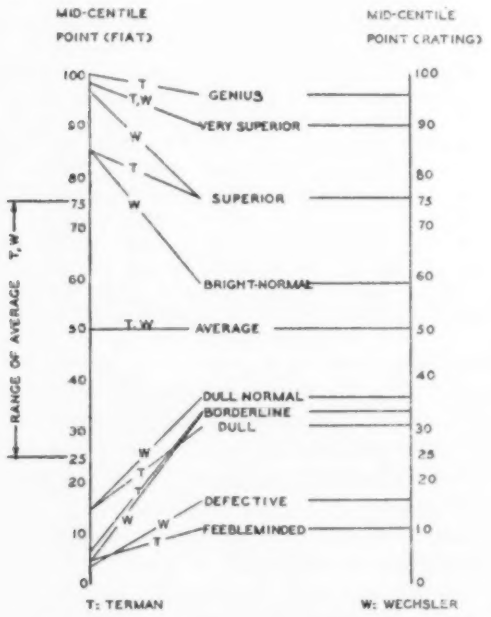


FIG. 1. Centile positions of various classification terms as suggested by Terman and by Wechsler and as found by empiric ratings.

If we attempt to create a new system for the classification of intelligence using our data, we are faced with problems 1 and 2 as stated above: how many classes to have, and where to place the limits of each class. We could use Wechsler's mathematical scheme, but this is unacceptable due to the too-wide range of "average." We could use Terman's multiples of IQ points, but this has no mathematical logic. One solution might be to use deciles, combining the two mid-deciles into one, and having nine "bands" of value. Using usual and the more homogeneous terms, and making some adjustments, we have Table 3. Although this scheme is not presented as ideal, on the basis of the information at hand it appears superior to Terman's and Wechsler's systems in that there is a closer reflection of what people actually take these terms to mean.

TABLE 3

A SCALE OF INTELLECTUAL CLASSIFICATION

| Decile | Mean Empiric Value | Classification | IQ Range |
|---|---|---|---|
| I | 90 | Very Superior | 121– |
| II | 77 | Very Bright | 114–120 |
| III | 76 | Superior | 109–113 |
| IV | 67 | High Average | 105–108 |
| V–VI | 50 | Average | 96–104 |
| VII | 35 | Low Average | 92– 95 |
| VIII | 27 | Inferior | 87– 91 |
| IX | 17 | Defective | 80– 86 |
| X | 11 | Feebleminded | – 79 |

This suggested scale places the two extreme classifications not at the middle of the deciles, but at the inner ends. No good term for the second decile appeared in our original 59 terms. "Very bright," although unacceptable both in connotation and in value, is the best compromise. Giving "average" a range of 20 centiles is much less generous than what has been suggested, but seems to reflect more adequately what our judges felt. For "feebleminded," we can only say that, even though experts place only 3 per cent of our population in this area, our judges were more generous.

If we are committed to the theory of modifying our classificatory system to fit current attitudes, rather than attempting a fiat system which makes arbitrary placements, a scale built in this fashion would seem to create less misunderstanding between psychologists and the public.

## SUMMARY

We have considered one instance of the problem of qualifying quantities, examined two standard systems of intellectual classification, and found certain discrepancies. The proposal that an international convention consider the matter of establishing a standard method is stated to be inferior to obtaining empirical data for the solution. By using a rating scale, mean values for 59 terms rated by five populations of ten people each, identified as intelligent adults, prison inmate teachers, psychologists, sociologists, and high school teachers were obtained. Eliminating awkward

and duplicating terms, we have 23 terms which are further considered. By accepting our rating scale as equivalent to a decile scale, we find certain gross differences between the placement of terms as suggested by Terman and Wechsler, and our empirical results. A rational scale of nine "bands" of values, using deciles, combining the fifth and sixth deciles, is suggested as a closer approximation to subjective impressions of test users.

*Received March 20, 1591.*

## References

1. Terman, L. M. *The measurement of intelligence.* Boston: Houghton Mifflin, 1916.
2. Wechsler, D. *The measurement of adult intelligence.* (3rd. Ed.) Baltimore: Williams & Wilkins, 1944.

# PSYCHOLOGICAL TESTS OF RESEARCH SCIENTISTS[1]

ANNE ROE
NEW YORK, N. Y.

THIS paper reports some special aspects of a research project which has been carried on for the past three and a half years. The major part of the project was the clinical study of eminent research scientists in various fields of science, and a consideration of differences and similarities among them in life histories and test performances. The full reports of the study have been published elsewhere [2, 3]. This paper presents certain interrelations between the tests used that are of particular interest. An earlier paper presented similar but less complete data on the biologists alone [4].

The tests employed were the Rorschach, the TAT, and a Verbal-Spatial-Mathematical (VSM) test constructed for the purpose by the Educational Testing Service, the three subtests being separately scored. The verbal test is an antonym test; the spatial test requires identification of drawings of the same object seen in different positions, and the mathematical test is chiefly one of mathematical reasoning. The mathematical test was not difficult enough for the physicists, so that they are not included in any correlations using it.

The subjects in the total study included 64 ranking research scientists: 20 biologists, 22 physical scientists, and 22 social scientists (psychologists and anthropologists). For three of these subjects there are no test data available, and there are some deficiencies in the data for the others, as is evident in the tables. These subjects were selected for their eminence in research, in the estimation of their peers. All are men, all but one are American-born, and all received the bulk of their training here and were still scientifically productive when studied. They range from 31 to 60 in age. Details on selection are given in the monographs, which also contain tables of raw data not reprinted here.

[1]Part of a study financed by a grant from the Division of Research Grants and Fellowships of the National Institutes of Health of the United States Public Health Service.

## EFFECT OF AGE ON TEST PERFORMANCE

Correlations with age are given for various aspects of test performance in Table 1.[2] The subtests of the VSM are entered first in the table. Clearly, only the spatial test shows any effect of age, within this age range, but this is great enough to be of some importance.

TABLE 1
EFFECT OF AGE ON TEST PERFORMANCE: CORRELATIONS

| Test Items | *N* | Correlation with Age |
|---|---|---|
| Verbal Test, number right | 59 | −.108 |
| Spatial Test, number right | 57 | −.397 |
| Mathematical Test, number right | 39 | +.001 |
| Rorschach: *R* | 61 | −.194 |
| *T/R* | 59 | −.150 |
| *W* | 61 | −.068 |
| *M* | 61 | −.204 |
| *ΣC* | 61 | −.204 |
| *F%* | 61 | +.113 |
| *F+%* | 61 | −.119 |
| *ITS* | 61 | +.222 |
| TAT: length of story | 61 | −.193 |

Not all the Rorschach determinants are appropriately handled by the usual correlational techniques, but a few which seem of particular interest, and which have wide ranges and reasonably normal distributions, have been included. None of these correlations is significant at the 5 per cent level ($r = .25$ for this sample).

The correlation between age and length of

[2]All calculations were done by Mr. Lassar Gotkin.

TAT story is also not significant. The measure of length is a crude one — the average number of lines of typescript for the stories for each man. These protocols were all recorded in shorthand; there are doubtless some omissions of words or phrases, but the record is reasonably accurate.

It would seem, then, that within this age range, and with the exception of the spatial test, variability in age does not have an important influence on test performance.

### VSM TEST INTERCORRELATIONS

These correlations, given in Table 2, are based on raw scores, i.e., on the number of items answered correctly in each subtest. Only the correlation between the verbal and spatial subtests is significant. This is not materially changed if age is partialled out. Descriptions of work on the spatial test would confirm that it can be done in different ways. This test was a rather unsatisfactory one, inasmuch as its significance is not clear. The distribution of scores was platykurtic, and its correlations with other tests (not used in this study) are confusing.

TABLE 2
VERBAL-SPATIAL-MATHEMATICAL TEST INTERCORRELATIONS: NUMBER RIGHT

| | Verbal | | Spatial | |
|---|---|---|---|---|
| | *N* | *r* | *N* | *r* |
| Spatial | 57 | +.332 | | |
| Mathematical | 39 | +.144 | 39 | +.212 |

### PRODUCTIVENESS OR OUTPUT

It can be hypothesized that there is a general factor of productiveness or ease and amount of output for the various tests, which might be manifested in such ways as the amount of work attempted in each test. For the VSM subtests these would be the number of items which the subject had tried to do (computed as the last one marked; if some items were omitted before that it was assumed that he had read and at least considered them), for the Rorschach the total number of responses, and on the TAT the average length of the stories. The correlations are given in Table 3. Two of the VSM correlations are low, but significant, but it is not certain that this is not primarily a speed factor inasmuch as this test was timed and some subjects did not finish all the items. But the correlation between verbal "tried" and Rorschach *R* is also significant, although low. Length of TAT story and number of responses on the Rorschach have a positive correlation of .499, which is the highest in the table.

### RELATIONS BETWEEN VSM AND PROJECTIVE TESTS

Relations between scores on the VSM and certain aspects of projective test performance which seemed relevant were studied by the correlations given in Table 4. Here are also given partial correlations which indicate the relative effects of age and total number of responses on the Rorschach (*R*). Total number of responses on the Rorschach clearly affects the total number of any particular kind of Rorschach response, and with the range in *R* as wide as it is in this group, this must always be considered. The mathematical test shows no significant relations with any of the other elements studied, but it is possible that there is some relation between this and percentage of Rorschach responses using Form alone (*F*%). (Correlation between *R* and *F*% in this group is −.028.) The group is small inas-

TABLE 3
INTERCORRELATIONS OF "OUTPUT"

| | S tried | | M tried | | *R* | | TAT length | |
|---|---|---|---|---|---|---|---|---|
| | *N* | *r* | *N* | *r* | *N* | *r* | *N* | *r* |
| Verbal, no. tried | 57 | +.317* | 39 | +.369* | 59 | +.264* | 59 | +.116 |
| Spatial, no. tried | | | 39 | +.225 | 57 | +.233 | 57 | +.095 |
| Mathematical, no. tried | | | | | 39 | +.098 | 39 | −.237 |
| Rorschach, *R* | | | | | | | 61 | +.499* |

*P = .05 or less.

TABLE 4

CORRELATIONS BETWEEN NUMBER RIGHT ON VSM SUBTESTS AND VARIOUS PROJECTIVE TEST MEASURES. PARTIAL CORRELATIONS WITH AGE AND *R*.

| 1 | 2. Verbal Right ($n = 59$) | | | 2. Spatial Right ($n = 57$) | | | 2. Math. Right ($n = 39$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | $r_{12}$ | $r_{12 \cdot age}$ | $r_{12 \cdot R}$ | $r_{12}$ | $r_{12 \cdot age}$ | $r_{12 \cdot R}$ | $r_{12}$ | $r_{12 \cdot age}$ | $r_{12 \cdot R}$ |
| Rorschach | | | | | | | | | |
| *R* | +.279* | +.265* | | +.048 | −.030 | | −.074 | −.075 | |
| *ITS* | −.007 | +.018 | −.110 | +.054 | +.159 | +.040 | −.213 | −.219 | −.204 |
| *W* | +.153 | +.146 | −.053 | +.062 | +.108 | +.040 | −.045 | −.046 | +.014 |
| *M* | +.221 | +.204 | +.322* | +.015 | −.073 | −.020 | −.137 | −.140 | −.133 |
| *F%* | −.003 | +.009 | +.005 | −.000 | +.049 | +.002 | +.234 | +.235 | +.263 |
| *F* + % | −.137 | −.152 | −.134 | +.340* | +.321* | +.337* | +.199 | +.201 | +.207 |
| Σ*C* | −.304* | −.337* | −.893* | −.025 | −.128 | −.103 | −.021 | −.022 | +.066 |
| TAT | | | | | | | | | |
| Length | −.545* | −.580* | | −.227 | −.254 | | +.145 | +.148 | |

*$P = .05$ or less.

much as the physicists did not take this test.

The spatial test shows significant relations only with the measure of the adequacy of form on the Rorschach (*F* + %), and age and *R* are apparently irrelevant.

The verbal test is positively correlated with total numbers of responses on the Rorschach and with number of human movement responses (*M*) when the effect of total responses is discounted. The negative correlation with sum of color responses (Σ*C*) is very high when *R* is partialled out. This is a rather consistent picture — the verbally skilled person, productive on the Rorschach, and rather introversive. The strong interest in persons shown also by the high *M* is also to be noted. I have elsewhere speculated that the reliance upon verbal modes of thinking and the high verbal skills demonstrated by individuals who are much concerned with personal relations may be due to the fact that words are our chief means of interpersonal communication, and those to whom this is important develop verbal facility. The negative relation with sum of color responses would indicate a much stronger degree of introversiveness than seems probable in this total picture. This is very likely due to the manner in which Σ*C* is computed. The *FC* responses are very common in those with high *M* [5, and shown also in my data], but they are weighted only half as strong as the *CF* responses; hence, the picture may be distorted. Interrelations within Rorschach elements themselves are not appropriately considered here, but they will be discussed at length in another paper.

The negative correlation between verbal right and TAT length is somewhat difficult to interpret. The correlation between verbal items tried and TAT length was negligible, although positive. Perhaps the point is that precision in the use of words is not related to fluency in their use.

### SHADING SHOCK AND VSM SCORES

The commonly accepted interpretation of shading shock as a pervasive, basic anxiety leads to speculation as to whether the presence of such anxiety has a disrupting effect upon intellectual performances of various sorts. If so, it might be manifested in lowered scores, or it might be manifested in greater scatter on test scores. A simple test is insufficient in so complicated a situation, but it seemed worth while to examine these data in this respect. The results are shown in Table 5, which gives

TABLE 5

SHADING SHOCK ON RORSCHACH AND VSM SCORES AND SCATTER

| VSM Scores | Shading Shock | | |
|---|---|---|---|
| | Strong $N = 16$ | Mild $N = 23$ | Absent $N = 22$ |
| Verbal right, ave. | 53.5 | 58.0 | 60.1 |
| Spatial right, ave. | 11.1 | 9.8 | 11.9 |
| Mathematical right, ave. | 12.9 | 14.0 | 16.8 |
| V-S, Scatter in σ scores | 1.0 | 1.1 | 0.7 |

mean *V*, *S*, and *M* scores and mean scatter for the subgroups of those with strong, mild, and no shading shock. The measure of scatter is

the difference between sigma score on verbal and on spatial tests. (The mathematical test was not taken by the physicists.) Although means for verbal and mathematical tests increase regularly with decreasing scatter, analysis of variance gives $F$'s of 1.00 and 1.20, which are not significant. On these data, then, it cannot be shown that shading shock has any effect on test score.

### MISCELLANEOUS COMPARISONS

A few further comparisons, which seemed of interest, have been made between Rorschach and TAT performance.

Length of TAT stories correlates + .555 with number of whole responses ($W$) on the Rorschach, which drops to + .339 when $R$ is partialled out, but is still significant. The point of this seems to be that the longer stories are also likely to be more complete, to encompass more, and this partakes of the same attitudes which result in large production of whole responses on the Rorschach.

One would also like to know whether originality on the two tests was related. For various reasons, it seemed preferable to use $P$ (popular responses) on the Rorschach, rather than $O$ (original responses). The measure on the TAT for each subject was the number of unusual stories he gave, plus one-half the number of stories to which he gave a distinctly unusual twist. In one instance, the measure is one of conformity, in the other, of departure from conformity. They should have a negative correlation if they have any basis in common. The correlation is completely insignificant, + .005.

Finally, Piotrowski's remark that the human movement responses ($M$) "indicate the testee's preoccupation with the future" stimulated another computation. The extent to which the subjects completed their TAT stories by including a future was extremely varied, even though they were specifically asked for this. To give stories with a future is not necessarily a measure of the extent to which the subject is preoccupied with the future, but it would seem probable that there is some relation. Therefore, a correlation was run between the number of human movement responses on the Rorschach and the number of stories which included the future. This is + .297, and, although low, is significant at the 2 per cent level.

### SUMMARY

This paper has reported some comparisons of test results on a verbal-spatial-mathematical test and the Rorshach and the TAT given to 61 eminent research scientists.

*Received March 20, 1951.*

### REFERENCES

1. PIOTROWSKI, Z. A Rorschach compendium—revised and enlarged. *Psychiat. Quart.*, 1950, **24**, 543–596.
2. ROE, ANNE. A psychological study of eminent biologists. *Psychol. Monogr.*, 1951, **65** (14), No. 331.
3. ANNE ROE. A psychological study of eminent physical scientists. *Genet. Psychol. Monogr.*, 1951, **43**, 121–235.
4. ROE, ANNE. Psychological examinations of eminent biologists. *J. consult. Psychol.*, 1949, **13**, 225–246.
5. WITTENBORN, J. R. Statistical tests of certain Rorschach assumptions: The internal consistency of scoring categories. *J. consult. Psychol.*, 1950, **14**, 1–19.

# SELF-EVALUATIVE QUESTIONNAIRES AS PROJECTIVE MEASURES OF PERSONALITY

GABRIEL ELIAS
THE UNIVERSITY OF ARKANSAS

BECAUSE of their recognized limitations, self-rating questionnaires are being rapidly replaced by projective devices in individual testing. Self-rating inventories are prevented by their very nature from measuring those layers of the personality which the testee is unwilling or unable to evaluate. Unfortunately, the various attempts to develop group projective devices [5, 8] have not produced techniques which are sufficiently practical to replace questionnaires for group testing as well.

The success of tests such as the Rorschach and the TAT stems largely from the fact that they give holistic interpretations of the personality Gestalt rather than isolated readings for individual traits. As a result, the services of a qualified clinician are required to interpret or "inspect" [10] the record, regardless of whether the test is given to groups or to individuals. The resulting expense is usually more then most testing programs can afford.

Because of the factor of expense, most attempts to improve group testing techniques have consisted of paper-and-pencil tests which purport, through some indirection or ingenuity, to elicit more honest answers than are obtainable by less subtle direct questioning [2]. These efforts will hold little promise so long as they continue to interpret self-ratings as direct measures rather than as mirrorings of the individual. No matter how sincere the testee and how honest his answers, the fact remains that his self-evaluations must be superficial and incomplete because they must necessarily confine themselves essentially to the areas of which he is consciously aware — and these we now know are often the least significant ones in an individual's personality [6]. Because projective devices interpret a testee's answers as behavioral reflections rather than as self-ratings of his personality [8], they often tell more about the individual than he knows himself. Much the same occurs in intelligence testing, in which the individual's test behavior is scored as a sample reflection of his capacity.

These considerations suggest that it may be possible to develop practical (easily administered and scored) instruments, that would prove more objective and penetrating than the current run-of-the-mill personality inventories, by scoring self-rating questionnaires as expressions rather than as ratings of personality. If, as is widely believed today, the individual is a "holistic" ("biosocial" or "organismic") indivisible whole, we would expect the testee, who is rating himself, to mirror his personality in addition to evaluating it.

If we apply these principles more specifically, we can predict that the answers a testee gives to a *true-false-?* type of self-rating health inventory would reflect a number of different personality characteristics. This study investigated the following three predicted relationships between the answers of testees on such inventories and their security-insecurity feelings at the time:

1. Insecure testees should answer *yes*, (i.e., complain of illness) in larger percentages than secure individuals. This hypothesis assumes that, since the individual is a biosocial whole, "psychological" inadequacies such as insecurity will be reflected in complaints of physical illness.

2. A larger percentage of insecure than of secure testees should select *?* answers. This hypothesis is supported partly by findings of other investigators that the tendency to select the *?* answer to such questions is a consistent personality characteristic [3] and by clinical evidence which considers disproportionate uncertainty a strong indicator of personal insecurity.

3. The more unstructured in (*a*) content, and (*b*) wording an item is, the greater the tendency of the insecure to exceed the percentage of secure testees who complain of illness. An unstructured question permits greater leeway for personality projection; therefore, it should reflect personality tendencies more sensitively than more structured items. The degree of structure of an item was determined by the extent to which its content and wording so controlled the inquiry situation as to eliminate sincere (i.e., unwitting) individual differences in its interpretation.

## PROCEDURE

An 112-item questionnaire devised by A. H. Maslow entitled "General Health and Hygiene Inventory" was given to several hundred Brooklyn College students. Several weeks later Maslow's Security-Insecurity (S-I) Inventory [5] was given without in any way relating the two tests.[1] For the purposes of this experiment, the Maslow S-I Inventory was regarded as a criterion of security-insecurity. The 122 individuals (27 per cent of the group) who scored most favorably on the S-I were taken as a "secure" criterion group, and were compared with an "insecure" criterion group of the 122 testees who made the most unfavorable scores.

The first 75 items of the health inventory were *true-false-?* questions. The remaining 37 were similar in content to related items of the true-false group, but were in different form. Item number 47, which is typical of the true-false questions, reads: "Do you consider yourself a rather nervous person?" The answer choices are *Yes, No,* and *?*. Item number 83, which is representative of the remaining 37 questions, is similar in content to item 47. It reads: "I have been nervous," and offers the following four answer choices, *Every day, Several times, One or two times, Not at all.* While this report is based on analysis of the results obtained for the first 75 items, the trends reported below appear to be confirmed by the results obtained for the remaining questions.

To verify part *a* of the third hypothesis (which relates the structure of item content to insecurity) the author compared the health inventory items of Group F with the items of Group E (See Table 1). The questions of Group F inquire about illnesses, such as scarlet fever and chicken pox, which we ordinarily ascribe to ourselves only after a physician has rendered a diagnosis. A definite and objective criterion, a medical diagnosis over which the testee has no control, determines whether he had the illnesses referred to in this group; the content of these questions is, therefore, highly structured. On the other hand, the content of the questions of Group E is clearly less structured; they refer to conditions, such as colds and insufficient weight, which the individual ordinarily diagnoses for himself without benefit of an outside criterion.

To investigate part *b* of the third hypothesis (relating structure of item wording to insecurity), the author studied 9 pairs of health inventory items. They were selected because in the opinion of the author the questions of each pair resembled each other closely in clinical content but differed sufficiently in wording structure to enable the author to designate one as the better structured item in each case. These pairings and structure evaluations (See Table 2) were tested on two different occasions. In 1946 the definition of item structure was typed on sheets on which the 18 questions were listed in random order. These sheets were given to 18 graduate students of psychology attending a seminar in attitudes conducted by H. H. Remmers at Purdue University with the request (*a*) to group the items into 9 pairs by matching each question with the one other item which resembled it most in clinical content, and (*b*) to designate the better structured question of each resulting pair. To check further on the accuracy of the writer's judgments, a similar study was undertaken at a meeting of faculty members and graduate students of psychology at the University of Rochester. After explaining the definition of structure used in this study, the author read the 18 items, two at a time, as he had originally paired them: after each pair was read, the audience was requested to indicate by a show of hands which was the better structured item.

The writer's opinions were confirmed by both studies. Fourteen of the 18 items were

[1] The writer is indebted to A. H. Maslow for administering these tests and making their results available to the writer.

paired in the same way as in Table 2 by 5 or more of the 18 Purdue judges, and, for each of these 14 pairs, a majority of the judges designated as better structured the one so designated in the table. The Rochester results were also in accord, for the majority of the group agreed with the author as to which item was better structured for 8 of the 9 pairs.

## RESULTS

All three hypotheses were confirmed.

The results summarized in Table 1 clearly confirm the first hypothesis. The 75 items of the health questionnaire are grouped into 8 categories according to similarity of clinical content to permit quick appraisal of the results. The critical ratios reported are for the differences between percentages of secure and of insecure who complained of illness. For category H, a minus sign precedes the lower end of the critical ratio range, because in the case of two items of that group the secure group alleged illness in greater percentages than the insecure. For no other item in any of the categories did this occur.

TABLE 1

ITEM CATEGORIES AND RANGES OF CRITICAL RATIOS

| Category | Clinical Content | CR Range* |
|---|---|---|
| A | "Psychological" Moods, e.g., "Do you often feel just miserable?" | 20.00 to 6.38 |
| B | Tiredness, Sleep, and Rest, e.g., "Do you almost always sleep well?" | 10.80 to 5.27 |
| C | Fear, e.g., "Have you ever been afraid of going insane?" | 8.02 to 1.51 |
| D | Evaluation of Own "Organic" Well-Being, e.g., "Do you usually feel well and strong?" | 10.00 to 1.56 |
| E | Illnesses and Ailments Which Are Usually Self-diagnosed, e.g., "Do you have colds frequently?" | 7.19 to 2.00 |
| F | Organic Illnesses, the Existence of Which Are Customarily Confirmed by Medical Diagnoses, e.g., "Did you ever have scarlet fever?" | 3.25 to 0.00 |
| G | Autonomic Rhythms, e.g., "Does your stomach get upset easily?" | 6.52 to 1.67 |
| H | Overt Symptom Behavior, e.g., "Did you ever have convulsions? | 0.00 to –0.50 |

*For all except two items of category H, the insecure answered "yes" (i.e., alleged illness) in as large or larger percentages than the secure.

For 71 of the 75 items the insecure professed illness in larger percentages than did the secure. Of these 71 items, 49 produced critical ratios of 3 or more; the ratios for 9 other items were 2 or more. Furthermore, of the 4 items that did not follow the trend of the 71, 2 were answered in identical proportions by both the secure and the insecure; and the other 2 (the 2 of Group H in which the insecure alleged illness in smaller percentages than the secure) produced critical ratios of less than 1.

The third hypothesis was supported with regard to both item content and wording structure. As appears in Table 1, the items of group E (whose content was poorly struc-

TABLE 2

COMPARISON BETWEEN ITEMS OF SIMILAR CONTENT BUT DIFFERENT IN STRUCTURE

| More Poorly Structured Item | | | Better Structured Item | | |
|---|---|---|---|---|---|
| Item No. | Statement of Item | *CR* | Item No. | Statement of Item | *CR* |
| 47 | Do you consider yourself a rather nervous person? | 14.78 | 7 | Can you sit still without fidgeting? | 11.4 |
| 70 | Do you usually feel well rested when you wake in the morning? | 10.80 | 62 | Do you usually feel fatigued when you wake in the morning? | 3.3 |
| 12 | Do you have colds frequently?" | 7.19 | 26 | Do colds sometimes lay you out for a long time? | 3.0 |
| 1 | Do you usually feel well and strong? | 5.91 | 5 | Do you have any outstanding physical defects? | 2.5 |
| 66 | Have you ever had spells of dizziness? | 5.63 | 72 | Do things sometimes swim or get misty before your eyes? | 0.4 |
| 20 | Do you get headaches frequently? | 4.37 | 67 | Do you have a great many bad headaches? | 3.0 |
| 21 | Do you get nauseous easily? | 4.02 | 22 | Do you faint easily? | 3.0 |
| 74 | Do your eyes often pain you? | 5.4 | 16 | Can you see well without glasses? | 3.4 |
| 23 | Do your bowels usually function well? | 4.52 | 60 | Are you troubled much by constipation? | 0.20 |

tured) differentiated much more clearly between the secure and the insecure than did the better structured ones of Group F. The range of ratios for Group E was 7.19 to 2.00, and for group F, 3.25 to 0.00. The results reported in Table 2 show that the same tendency holds for differences in wording structure. For each of the 9 pairs, the question which the author (and both the Purdue and Rochester judges) judged to be poorly structured, differentiated more sensitively between the secure and the insecure than did the better stuctured item. Apparently, the less structured the content or wording of the questions, the more did testees' answers express differences in security-insecurity feelings.

## DISCUSSION

Practical and theoretical considerations dovetail in the present study. The writer's initial concern was a very practical one: Why, in view of the recognized inadequacies of personality inventories, have they remained so popular for group testing? Why have we failed to develop more acceptable tools?

The answer appears to lie in the lag between our theoretical postulates and their application. Practically every psychologist and psychiatrist, be he a behaviorist or field theorist, a Freudian or an Adlerian, accepts the "holistic," "organismic," or "biosocial" concept of personality. The individual behaves as a whole. He expresses his personality in his behavior. Yet, rating questionnaires must, perforce, violate this principle. When, as in opinion polling, we seek the most objective answer possible, we usually try to word our questions and structure our test situations so as to eliminate all room for projection of personality [3]. We hope thus to avoid subjectivity and irrelevant personality reflections. But is this possible? We know that testees lie, repress, and distort when rating themselves. A self-rating, when it is valid, is accurate only because the testee's entire personality is such that he can properly evaluate himself under the circumstances. Because a self-rating, accurate or inaccurate, is behavior and an expression of the entire individual, its very accuracy or inaccuracy will depend on his personality.

The findings of the present study indicate that even when answering a simple question regarding his health, an individual projects feelings of security or insecurity. They show, moreover, that personality projects itself in such instances in predictable fashion, and, therefore, that much can be gained by constructing paper-and-pencil techniques which apply rather than flaunt holistic theoretical considerations. We would do better to design tests so as to predict personality from projective behavior than to seek to devitalize a test in a fruitless attempt to prevent it from eliciting personality projection.

The writer has already completed one rather extensive study based on these principles [5] in his construction of the Elias Family Opinion Survey. This test measures personal feelings of homey-homelessness projectively. A testee is asked to express his opinions regarding "the general family life that existed in your community when you were a child." A typical question is 60, which reads, "Children love one of their parents in ways that aren't nice." In the process of expressing his attitudes regarding general family life, he projects his own feelings, and these reflections are as objectively and easily scored as are the self-ratings of typical attitude and personality inventories. Various studies with the test prove it to be reliable and valid [5, 12].

In another application of this study's findings, one of the author's students gave the adult form of the California Test of Personality and 32 selected items of Maslow's General Health and Hygiene Inventory to 82 University of Arkansas psychology students. He found that a significantly larger percentage of those whose scores fell into the most maladjusted quarter on the California test gave four or more *?* answers to the health inventory than did those who scored in the most adjusted quarter ($P$ [$\chi^2$] $< .05$). This finding can be extended to attitude testing and any other area where *T-F-?* questions are used. Are the results of these tests somewhat distorted by personality factors such as the stronger tendency of the more insecure individuals to select the *?* answer to questions irrespective of their content?

## SUMMARY AND CONCLUSIONS

Despite the admitted inadequacies of self-rating questionnaires, they are by far the most

used instruments for group testing. Theoretical considerations suggest that self-rating questionnaires can be improved in validity and effect by interpreting the test answers as reflections, rather than as direct measures, of personality.

To ascertain whether predictable personality projection can be elicited in this manner, Maslow's security-insecurity inventory and a self-rating *T-F-?* type of health questionnaire were given to Brooklyn College students. When comparing the 27 per cent most insecure (as measured by Maslow's security-insecurity inventory) with the 27 per cent most secure, the following predicted relationships were confirmed:

1. The insecure testees tended to rate themselves as being ill and to select the *?* answer in greater percentages than did the secure testees.

2. The health inventory items which were unstructured (both in content and wording) differentiated more clearly between the secure and the insecure than did other items of better structure despite the essential similarity of clinical content of the two groups of items.

These findings and the theoretical considerations from which they are derived suggest a fruitful area for test-builders. Objective, easily scorable tests can be constructed by interpreting self-ratings as behavioral mirrorings rather than as direct, evaluating measures of personality.

*Received March 26, 1951.*

## References


1. Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.
2. Campbell, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.
3. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.* 1950, **10**, 3–21.
4. Elias, G. Construction of a test of non-homeyness and related variables. Unpublished Doctor's thesis, Purdue Univ., 1949.
5. Elias, G. The Elias Family Opinion Survey, a measure of homelessness. *J. abnorm. soc. Psychol.*, in press.
6. Freud, S. Repression. In *Collected papers.* London: Hogarth, 1925. Vol. IV, pp. 84–97.
7. Harrower-Erickson, Molly R. Developments of the Rorschach test for large scale application, *Rorschach Res. Exch.*, 1944, **8**, 125–140.
8. Maslow, A. H., *et al.* A clinically derived test for measuring psychological security-insecurity, *J. gen. Psychol.*, 1945, **33**, 21–41.
9. Munroe, Ruth L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method, *Appl. Psychol. Monogr.*, 1945, No. 7.
10. Murphy, G. *Personality.* New York: Harper, 1947.
11. Murray, H. A. *Explorations in personality.* New York: Oxford Univ. Press, 1938.
12. Richmond, Elise. Relationship between selected variables and homelessness test scores, and suggested varying standardizations. Unpublished Master's thesis, Univ. of Arkansas, 1950.
13. Wang, C. K. A. Suggested criteria for writing attitude statements. *J. soc. Psychol.*, 1932, **3**, 367–373.

# A PSYCHOMETRIC STUDY OF DRINKERS[1]

PETER J. HAMPTON
WESTERN RESERVE UNIVERSITY

THE basic purpose of this study was (1) to compare a group of alcoholics with a group of comparable nonalcoholics on a number of tests; (2) to select, on the basis of the comparison, those test items with respect to which statistically significant differences between alcoholics and nonalcoholics were found; (3) to construct from these items a reliable and valid personality questionnaire for drinkers which might differentiate not only alcoholics from nonalcoholics, but among different categories of alcoholics and nonalcoholics.

## PROCEDURE AND RESULTS

A battery of tests consisting of a Personal History Questionnaire, the Otis Self-Administering Tests of Mental Ability, the Strong Vocational Interest Blank for Men, and the Minnesota Multiphasic Personality Inventory was administered to a group of 84 male alcoholics and a comparable group of 84 male nonalcoholics. Criteria used for inclusion of subjects in the experimental group of the test tryout part of the study were as follows: (1) membership in Alcoholics Anonymous, (2) pronouncement by one or more psychiatrists that the subject was an alcoholic, and (3) hospitalization or commitment to jail for alcoholism. The criteria used for inclusion of subjects in the control group of the test tryout part of the study were: (1) inapplicability to control subjects of the three criteria used for the selection of experimental subjects, (2) subject's own statement that he was not an alcoholic, and (3) close approximation in age, intelligence, education, and socioeconomic level of control subjects to experimental subjects.

Test results for alcoholics and for nonalcoholics were reported and expressed in means, standard deviations, and *t*-ratios. Tests revealing statistically significant differences in mean scores between alcoholics and nonalcoholics were submitted to further study. An item analysis of the Minnesota Multiphasic Personality Inventory and parts of the Personal History Questionnaire revealed a fund of items with respect to which statistically significant differences were found between alcoholics and nonalcoholics. For most of the responses to these items the differences between alcoholics and nonalcoholics were of a magnitude that would occur less than once in 100 times as the result of sampling errors; in nearly all of them, the differences would occur no more than 5 times in 100 as a result of sampling errors.

Discarding all the items with regard to which a critical ratio of less than 2.00 was obtained left 125 items from the Minnesota Multiphasic Personality Inventory and 31 items from the Personal History Questionnaire. Responses to these items were weighted in accordance with Strong's Tables of Net Weights for Differences in Per Cents, and the items were included in a trial Personality Questionnaire for Drinkers. Part 1 of the study thus resulted in the development of a trial questionnaire which differentiated alcoholics from nonalcoholics. The next step was to find out whether this trial questionnaire would differentiate other alcoholic and nonalcoholic groups.

In part 2 of the study a new sample of alcoholics and nonalcoholics was used. The trial Personality Questionnaire for Drinkers was administered to 250 male subjects. Altogether, 100 male alcoholics, separated into groups of 50 secondary alcoholics and 50 week-end alco-

[1]From a thesis submitted to the Graduate Faculty of Western Reserve University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This report was read at the Annual Meeting of the Canadian Psychological Association, Toronto, May 1950.

holics, were compared with 150 comparable male nonalcoholics, separated into groups of 50 social family drinkers, and 50 social occasion drinkers, and 50 near or total abstainers. Inclusion in the various groups was by self-rating.

Test results for alcoholics and nonalcoholics, and for each subgroup, were again recorded and expressed in means, standard deviations, and *t*-ratios. Significant differences were found not only between the mean scores of the alcoholic and nonalcoholic groups, but also between secondary alcoholics and week-end alcoholics, social family drinkers, social occasion drinkers, and near or total abstainers; between week-end alcoholics and social family drinkers, social occasion drinkers, and near or total abstainers; and between social family drinkers and social occasion drinkers and near or total abstainers.

Test norms, validity, and reliability data for the trial Personality Questionnaire for Drinkers were obtained from the check sample population. The validity of the trial Personality Questionnaire for Drinkers was investigated by the extreme-groups method and the expectancy table technique; the reliability of the trial Personality Questionnaire for Drinkers was checked by the split-half method and the test-retest method. Both the validity and reliability (test-retest coefficient of reliability = .87; corrected split-half coefficient of reliability = .89) of the test were found to be acceptable. Finally, a tentative cutting score of 59 on the Questionnaire, dividing alcoholics or potential alcoholics from nonalcoholics, was arrived at by inspection of ogive curves.

## CONCLUSIONS AND IMPLICATIONS

It has been demonstrated in this study that significant differences in responses to items from a Personal History Questionnaire and the Minnesota Multiphasic Personality Inventory exist between alcoholics and nonalcoholics of similar backgrounds. This finding supports the hypothesis that alcoholics can be adequately differentiated from nonalcoholics on the basis of differences in responses to certain items on the Personal History Questionnaire and the Minnesota Multiphasic Personality Inventory.

It has further been shown that the trial Personality Questionnaire for Drinkers, constructed with items from the Personal History Questionnaire and the Minnesota Multiphasic Personality Inventory with respect to which statistically significant differences between alcoholics and nonalcoholics were found, when administered to new samples of alcoholics and nonalcoholics, differentiates not only alcoholics from nonalcoholics, but different categories of alcoholics and different categories of nonalcoholics from each other. In particular the Personality Questionnaire for Drinkers has been shown to differentiate (*a*) alcoholics from nonalcoholics, (*b*) secondary alcoholics from week-end alcoholics, social family drinkers, social occasion drinkers, and near or total abstainers, (*c*) week-end alcoholics from social family drinkers, social occasion drinkers, and near or total abstainers, and (*d*) social family drinkers from social occasion drinkers and near or total abstainers.

The Personality Questionnaire for Drinkers is believed to have sufficient validity and reliability to be of value as a screening technique for drinkers. It may also be of value as a guide in vocational guidance, employee selection, and clinical diagnosis.

Results of this study show that mean scores for different groups of drinkers on the Personality Questionnaire for Drinkers rise in magnitude as alcohol increases in importance in the life pattern of the variously classified groups. The more maladjusted the individual is on the average, the more need he seems to show for alcohol as a crutch. The results, in so far as they are representative of other population samples, therefore may indicate that individuals who show high scores on the Personality Questionnaire for Drinkers, irrespective of present status of drinking, may be more apt to use alcohol as an escape under extreme emotional stress. The Personality Questionnaire for Drinkers will, of course, have to be used with caution until further research corroborates our findings.

Results of this study also suggest that whereas near or total abstainers find little or no use for alcohol in their life adjustments, they show, on the average, fewer favorable responses on the Personality Questionnaire for Drinkers than do social drinkers. On first im-

pression this suggestion is startling. Upon examination, however, the facts are not so disparate after all. Near abstainers and especially total abstainers may also depend upon the concept of "drinking" as a crutch — but in a different way. Drinkers seem to depend upon drinking largely as a defense mechanism; abstainers may depend upon not-drinking as a defense mechanism. This explanation is suggested on the basis of responses made by total abstainers to item number 97 on the Personality Questionnaire for Drinkers, which reads, "I believe that a person should never taste an alcoholic drink." The majority of total abstainers answered this statement by encircling "True." In addition, a number of total abstainers amplified their feelings about drinking by such write-ins as, "I have never touched liquor and I hope I never will!" and "I would rather be dead than touch the stuff!"

Finally, no uniform personality pattern was found on the Personality Questionnaire for Drinkers which was characteristic of all intemperate drinkers. There were several patterns, however, which because of their incidence may be presumed to characterize large numbers of the statistical population of alcoholics. Analysis of items on the Personality Questionnaire for Drinkers indicative of disabling psychological abnormality suggests the presence of the following clinically defined constellations in the personalities of alcoholics: In order of occurrence they are: neurotic, psychopathic, compulsive, hypomanic, schizothymic, and homoerotic.

## SUGGESTIONS FOR FURTHER RESEARCH

The statistical treatment of data and the conclusions discussed in this study should not be considered final statements. The construction of screening techniques for psychometrically differentiating primary alcoholics from secondary alcoholics, secondary alcoholics from week-end alcoholics, week-end alcoholics from social family drinkers, social family drinkers from social occasion drinkers, and all of these from total abstainers, is a recent development. Extensive research will be necessary if such screening techniques as the Personality Questionnaire for Drinkers are to be used wisely in diagnosis, prediction, and control of "drinking" behavior. The following suggestions are made to encourage further research on screening techniques for drinkers.

1. *Search for significant items.* While the present study has located a substantial number of test items which significantly differentiate different categories of drinkers from each other, the search for such items should be continued both by way of try-out of already existing test items and by construction of new items.

2. *Need for validation studies.* The adequacy of items for differentiating different categories of drinkers established by way of critical ratio procedures should be checked by more refined statistical techniques. There is need for correlational studies. Analysis of variance and factorial analysis would be of advantage. There is need, also, for the establishment of occupational norms and critical scores for different groups of drinkers.

3. *Personality characteristics of different types of drinkers.* The present study has isolated several constellations of personality among intemperate drinkers. Correlational studies determining the degree of relationship between responses to items of the Personality Questionnaire for Drinkers indicative of different personality patterns, and responses to items of different tests purporting to measure these patterns, would provide a clearer picture of the amount of relationship between these patterns in the personality of different types of drinkers.

4. *Stability of personality characteristics of different types of drinkers.* In addition to locating personality constellations among drinkers there is also need for studying their stability. This can be done on a longitudinal basis by retesting periodically to identify recurring patterns or shifts to different patterns, or on a cross-sectional basis by comparing test results for different age levels.

5. *Moderate drinking and personality adjustment.* Since extent and intensity of intemperate drinking seem to be related to personality maladjustment, it may well be that moderate drinking is related to personality adjustment. This possibility should be investigated.

6. *Personality constellations of abstainers.* An adequate study of "pure" abstainers should be made to see to what extent the compulsive

nature of their personality, suggested by this study, is related to intemperate drinking, moderate drinking, and no drinking.

7. *Occupational rating scale for drinkers.* It might also be of importance to vocational psychologists to have competent judges, who are familiar with the duties of a wide range of jobs, rate a list of representative occupations on a five-point scale for extent of drinking.

*Received February 16, 1951.*

## References

1. Hampton, Peter J. The development of a personality questionnaire for drinkers. Unpublished Doctor's thesis, Western Reserve Univ., 1950.
2. Manson, M. P. A psychometric analysis of the psychoneurotic and psychosomatic characteristics of alcoholics. *J. clin. Psychol.*, 1949, **5**, 77–83.
3. Manson, M. P. A psychometric anaylsis of psychopathic characteristics of alcoholics. *J. consult. Psychol.*, 1949, **13**, 111–118.
4. Seliger, R. V. Are you an alcoholic? *Alcohol. Hyg.*, 1946, **2**, 5–10.
5. Tiebout, H. M. The syndrome of alcohol addiction. *Quart. J. Stud. Alcohol*, 1945, **5**, 535–546.
6. Wittman, Phyllis. A controlled study of the development and personality characteristics of chronic alcoholics. *Elgin Papers*, 1939, **3**, 77–84.
7. Wortis, H., Sillman, L. R., and Halpern, Florence. *Studies of compulsive drinkers.* New Haven: Hillhouse Press, 1946.

# PRACTICAL CORRELATES OF PSYCHIATRIC SYMPTOMS

J. R. WITTENBORN[1]
YALE UNIVERSITY
AND
FRED A. METTLER
COLUMBIA UNIVERSITY

THERE is evidence that the symptoms manifested by mental hospital patients cluster together in a systematic fashion [2, 3]. The nature of these clusters has been determined through factor analyses of symptom rating scales. The rating scales were developed in collaboration with psychiatrists and presumably are a good and representative sample of the types of currently discernible symptoms which psychiatrists consider important in making a descriptive diagnosis of mental hospital patients whose disorder could be described as functional. They are primarily descriptive in nature and make no particular demand upon the theoretical sophistication of the rater or his skill in making psychoanalytic inferences. Although the scales should not be considered representative of all the various kinds of psychiatric disorders which may be encountered, the nature of the symptom clusters which are found among these rating scales is plausible from the standpoint of the descriptive concepts employed in the diagnosis of psychiatric patients.

The factor analysis studies provide the basis for a quantitative weighting of groups of symptom rating scales so that it is now possible to rate the symptom manifestations of a mental hospital patient and to score these symptom ratings with respect to each of nine different symptom clusters; i.e., it is possible to describe a given patient quantitatively with respect to nine different empirically determined symptom clusters. Tentative labels have been offered for these symptom clusters, and the general nature of each symptom cluster may be determined by referral to the original studies [1, 2, 3], which show the precise degree to which each symptom rating scale participates in the determination of each of the various clusters.

Obviously, there are many aspects of the patient's existence which are a part of, or are related to, the general nature of his disorder but which are not likely to be included in psychiatric symptom rating scales, particularly in rating scales which are restricted to currently discernible symptom manifestations. Patients' scores on the various symptom clusters have been correlated with scores on the Wechsler-Bellevue subtests [4] and also with various Rorschach scores [5]. These correlations contribute little to our understanding of the general implications of the symptom clusters and, incidentally, contribute less to our confidence concerning the possibility of using scores from either of those tests as an aid in a descriptive diagnosis.

In a recent study [6], the symptom cluster scores have been correlated with biographical factors found in the patient's hospital records. This study shows that among hospitalized patients there are significant relationships between the patient's score on the symptom clusters and characteristics of the onset of the patient's disease. Similarly, biographical situations or characteristics which existed before the patient was recognized as ill were shown to be related to current symptom clusters ob-

[1]This investigation as supported by a research grant (MH 118C, Dr. Fred A. Mettler responsible investigator) from the National Institute of Mental Health, of the National Institutes of Mental Health, Public Health Service. The writers also acknowledge the cooperation of the State of New Jersey and the staff at Greystone Park.

TABLE 1

CHI SQUARES FOR TESTS OF INDEPENDENCE BETWEEN OCCUPATIONAL SCALE AND SYMPTOM CLUSTER SCORE

| Scale | Manic-Depressive, Manic | Schizophrenic Excitement | Acute Anxiety | Manic-Depressive, Depressed | Paranoid Schizophrenia | Paranoid Condition | Conversion Hysteria | Phobic-Compulsive | Hebephrenic or Deteriorated Schizophrenia |
|---|---|---|---|---|---|---|---|---|---|
| 1. Incapable of carrying out tasks | | 9.1 | −5.3* | | | | | | |
| 2. Extremely disagreeable | | | | | | | | | |
| 3. Takes pride in quality of work | | | 5.3 | | | | | | |
| 4. Not reassured of doing worth-while work | | | | | | | | | |
| 5. Acutely uncomfortable | | | | | | | | | |
| 6. Interest aroused with difficulty | | 5.6 | | | | | | | |
| 7. Difficult to draw interest to a task | | | | | | | | | |
| 8. Insistently argumentative | | | | −7.0 | | | | | |
| 9. Insists on continuous reassurance | | | | | | | | | |
| 10. Spontaneously assaultive | | | | | | | | | |
| 11. Ideas consistently bizarre | | | | | | | | | |
| 12. Consistently careless worker | | | −4.8 | | | | | | |
| 13. Characteristically hilarious | | | | | | | | | |
| 14. Oblivious to demands of task | | 5.6 | −4.8 | | | | | | |
| 15. Unrealistically overconfident | | | | −5.9 | | | | | |
| 16. Fails to comprehend requirements of situation | | | −4.8 | | | | | | |
| 17. Grossly inconsiderate | 6.3 | | | | | | | | |
| 18. Habitually hypercritical | | | | | | | | | |
| 19. Despondent, dejected, retarded | | | | | | | | | |
| 20. Oblivious of feelings of others | | 8.4 | | 8.9 | | | | | 5.2 |
| 21. Unable to attend to activities | | | −4.8 | | | | | | |
| 22. Difficulty understanding problems | | 8.4 | −6.1 | 5.8 | | | | | 5.2 |
| 23. Feels people are against him | | | | −5.9 | | | | | −3.8 |
| 24. Openly hostile and destructive | | | | | | | | | |
| 25. Beliefs unrealistic and delusional | | | | | | | | | −4.2 |
| 26. Behavior hurried, impulsive | 5.9 | | | | | | | | |
| 27. Completely lacking in initiative | | | | | | | | | |
| 28. Inappropriate speech and actions | 8.9 | | −4.7 | | | | | | |
| 29. Violates rules and regulations | | 8.9 | | | | | | | 8.5 |
| 30. Serious lack of drive | | | | | | | | | |
| 31. Manneristic behavior | | | | | | | | | |
| 32. Complete lack of persistence | | | | | | | | | |
| 33. Generally unkempt | | | | | | | | | |
| 34. Expends excessive effort needlessly | | | | | | | | | |
| 35. Fastidious concern for next phase of work | | | | −6.2 | | | | 8.6 | |
| 36. Lacks stamina | | | | | | | | | |
| 37. Cannot concentrate | 6.2 | | | | | | | | |
| 38. Reluctant when induced to begin work | | | | | | | | | |
| 39. Verbalizations characterized by profanity | | | | | | | | | |
| 40. Gross lack of dexterity | | 5.5 | | 5.9 | | | | | |
| 41. Reacts strongly to criticism | | | | −5.9 | | | | | |
| 42. Demoralized by failure | | | | | | | | | |
| 43. Resentful of task | | | | | | | | | |
| 44. Continuous speech or actions | 6.5 | | | | | | | | |
| 45. Responds only when approached or detained | | | | | | | | | |
| 46. Markedly overmeticulous and precise | | | | | | | | | |
| 47. Makes critical and unkind remarks | | | | −18.8 | | 5.4 | | | |
| 48. Lacks awareness of errors | 5.7 | 12.1 | | | | | | | |
| 49. Selfish and uncooperative with others | 3.8 | | | | | | | | |

Table 1 (continued)

| Scale | Manic-Depressive, Manic | Schizophrenic Excitement | Acute Anxiety | Manic-Depressive, Depressed | Paranoid Schizophrenia | Paranoid Condition | Conversion Hysteria | Phobic-Compulsive | Hebephrenic or Deteriorated Schizophrenia |
|---|---|---|---|---|---|---|---|---|---|
| 50. Interests changeable | | | | | | | | | |
| 51. Laughs inappropriately | | | | | | | | | |
| 52. Speech usually terse | | | | 6.2 | | | | | 5.6 |
| 53. Works at extremely slow pace | | | | | | | | | |
| 54. Extremely submissive | | | | | | | | | |
| 55. Suggestions fantastically inappropriate | | | −4.8 | | | | | | |
| 56. Exaggerated suspicions of others | | | | | | | | | |
| 57. Efforts characterized by sporadic fluctuations | | | −5.3 | | | | | | |
| 58. Unconcerned about his behavior | | | | | | | | | |
| 59. Negativistic and contrary | | | | | | | | | |
| 60. Refuses to participate in activities | | | | | | | | | |
| 61. Believes activity aggravates symptoms | | | | −4.0 | | | | | |

*Negative sign for chi square indicates an inverse relation between occupational scale and cluster score.

tained from ratings made at the time the patients were admitted to the hospital. The types of correspondence observed in this study are of interest not only because they broaden our concept of the general behavioral significance of various types of symptom manifestations, but also because they encourage us in the study of the biographical origins of mental illness.

The symptom cluster scores have been correlated with the diagnosis actually given the patient by psychiatrists [7]. This investigation is of considerable interest because it has shown that the labels given symptom cluster scores are appropriate in the sense that the various symptom cluster scores are most highly correlated with the psychiatric diagnosis which customarily bears the label assigned to the symptom cluster score.

## THE PRESENT ANALYSIS

In order to extend our knowledge of the general behavioral significance of symptom cluster scores, socially relevant aspects of the patient's behavior should also be studied. Accordingly, the present study is designed to relate the kinds of socially practical behavior observed in occupational therapy situations (and the other special services of the hospital) with the symptomatic behavior represented by the symptom cluster scores. The present report describes a statistical investigation in which each of 61 rating scales appropriate for occupational therapy situations is correlated with each of the nine symptom cluster scores.[2] The patients are a group of 60 schizophrenics who were hospitalized at Greystone Park, New Jersey, for a total of at least three years, and who were considered as possible candidates for psychosurgery. The set of occupational rating scales was developed by the staff of the Northampton Veterans Administration Hospital in collaboration with the writer. A typical item from this rating scale would read in the following manner:

1. Shows average or better ability to understand directions.
2. Understands adequately with examples and repeated instructions.
3. Capable of understanding only the routine aspects of a problem.
4. Has extreme difficulty in understanding the most routine aspects of a problem even after repeated directions.

In all cases high ratings for the occupational scales indicate an extreme of the behavior sampled by the scale. Similarly, high symptom cluster scores indicate a high degree of the type of symptoms indicated by the cluster.

The results of the analysis are shown in

[2]The occupational ratings were made by occupational therapists, attendants, or other nonmedical supervisors. The psychiatric symptom ratings were made independently by psychiatrists.

Table 1, which includes a terse phrase from the most extreme statement of each of the rating scales. Table 1 shows $\chi^2$ statistics which are based on fourfold tables and which relate the patient's status on each of the rating scales to his status on each of the symptom cluster scores. Only the $\chi^2$ statistics sufficiently large to be significant at the 5 per cent level are included in Table 1. Of the 549 possible relations, 45 of them are significant at the 5 per cent level. This is approximately twice as many as would be expected if chance alone were determining the size of the $\chi^2$ statistic. Inasmuch as the analysis was conducted to broaden our understanding of the behavioral significance of the various symptom cluster scores, a discussion of the results of Table 1 will proceed from symptom cluster to symptom cluster.

The first column in the table is for the manic-state cluster score. There is a tendency for high scores on this cluster to be associated with item 17, grossly inconsiderate; 26, hurried, impulsive behavior; 28, inappropriate speech and actions; 37, cannot concentrate; 44, spontaneous and continuous speech or action; 48, lack of awareness of errors; and 49, selfish and uncooperative with others.

It is interesting to observe that the excited cluster score is quite different from the manic cluster score. Actually, the two cluster scores have but one scale in common, number 48—both the excited and the manic patients seem unaware of their errors. Patients who tend to have a high score on the excited cluster are characterized in this way: 1, incapable of carrying out tasks; 6, interest is aroused with difficulty; 14, are oblivious to the demands of a task; 20, are oblivious to the feelings of others; 22, have difficulty understanding problems; 29, violates rules and regulations; 40, there is a lack of dexterity; 48, are unaware of their errors.

The patients who stand high on the anxiety cluster score tend to have the following kinds of scale ratings: 1, they are not incapable of carrying out assigned tasks; 3, they take pride in the quality of their work; 12, they are not consistently careless workers; 14, they are not oblivious of the demands of a task; 16, they are not the ones who fail to comprehend the requirements of a situation; 21, they are not the ones unable to attend to activities; 22, they are not the ones who have difficulty understanding a problem; 28, they are not characterized by speech and actions inappropriate to the situation; 55, they do not give fantastically inappropriate suggestions; and 57, their efforts are not characterized by sporadic fluctuations. In general, it could be inferred that in this sample of schizophrenic patients, those who had the highest ratings on anxiety were the patients who in general were free from many kinds of occupational disabilities. The patients who stand high on the anxiety cluster do not tend to be rated in a manner similar to the patients who stand high on the manic cluster or on the excited cluster. As a matter of fact, the patients with high anxiety scores and high excitement scores tend to be characterized in opposite respects by two occupational rating scales.

Patients who stand high on the depressed cluster score tend to be characterized in the following respects: 8, they are not insistently argumentative; 15, they are not unrealistically overconfident; 20, they are oblivious of the feelings of others; 22, they do have difficulty understanding problems; 23, they don't tend to feel that people are against them; 35, they seem not especially concerned about the next phase of a task; 40, there is a lack of dexterity; 41, they do not react strongly to criticism; 47, they do not make critical remarks about others; 52, their speech is terse; 61, they do not believe that activity is going to aggravate their symptoms.

It is most interesting to observe that the symptoms of the paranoid schizophrenics as measured by the symptom cluster score bear no relation to any of the 61 items sampled by the occupational rating scales. In view of the numerous plausible respects in which high scores on the other symptom clusters are related to various aspects of the occupational rating scales, this is a conspicuous finding. This suggests that for the most part, the disturbance of the paranoid schizophrenic is primarily related to the symptoms which are of interest to the psychiatrist, and that the disturbance of paranoid schizophrenia does not detract to any particular degree from the patient's performance on the kinds of occupational activities in which institutional patients en-

gage. It is commonly reported by institutional workers that the paranoid patients are among the most useful around the hospital. It is probably because of this fact that many paranoid patients are released as competent when in effect they are still seriously ill and although possibly occupationally competent, grossly incompetent from the standpoint of social and interpersonal existence.

The lack of relation between occupational ratings and scores on both the paranoid schizophrenic cluster and the paranoid-condition cluster is quite striking and suggests that the paranoid patient, whether he be considered schizophrenic or not, is generally a patient who adapts reasonably well to the requirements of simple occupational activities.[3]

The paranoid-condition cluster score is not especially characteristic of the present sample and is related to only one rating scale, and this one has more social significance than significance for competence in a simple occupation. For scale number 47, it is shown that these patients tend to make critical and unkindly remarks about others.

TABLE 2

DISTRIBUTION OF SYMPTOM CLUSTER SCORES FOR THE SAMPLE OF 60 SCHIZOPHRENICS

| Symptom Cluster | Scores 1 and 2 | Scores 3 and 4 | Scores 5 and 6 | Scores 7 and 8 | Scores 9 and 10 |
|---|---|---|---|---|---|
| Manic | 17 | 8 | 7 | 6 | 2 |
| Schizophrenic Excitement | 9 | 11 | 8 | 9 | 3 |
| Acute Anxiety | 30 | 5 | 2 | 3 | 0 |
| Depressed | 8 | 7 | 9 | 8 | 8 |
| Paranoid Schizophrenia | 6 | 17 | 7 | 5 | 5 |
| Paranoid Condition | 22 | 13 | 3 | 1 | 1 |
| Conversion Hysteria | 36 | 2 | 1 | 1 | 0 |
| Phobic-Compulsive | 29 | 7 | 1 | 2 | 1 |
| Hebephrenic Schizophrenia | 14 | 8 | 8 | 7 | 3 |

Scores on the conversion hysteria cluster also are seen to bear no relation to occupational aspects of the patient's behavior. This, however, merely reflects the fact that almost no patients in the present sample show strong conversion tendencies.

It is noted that patients with phobic-compulsive symptoms are relatively rare in the present sample and are not particularly characterized by any of the occupational scales. It seems plausible that patients with phobic-compulsive symptoms, who are placed in any tolerable activity, would tend to work without gross handicap. There is one respect in which the phobic-compulsive patient is characterized by the occupational scales. Specifically, item 35 shows him to be fastidiously concerned about the next phase of his work.

Patients who have high scores on the hebephrenic or deteriorated type of schizophrenia are characterized in certain respects by their occupational ratings. For example, 20, these patients are oblivious of the feelings of others; 22, they have difficulty understanding problems; 23, they are not inclined to feel others are against them; 25, they do not have grandiose or grossly unreasonable beliefs in their ability; 29, they consistently violate rules and regulations; and 52, their speech is usually terse. The patients with the high scores on the hebephrenic-type cluster are seen to resemble patients who have high scores on some of the other symptom clusters. For example, items 20 and 22, being oblivious of the feelings of others and having difficulty in understanding problems, are characteristic of patients with high scores on the excited cluster, those who have high scores on the depressed cluster, and those with high scores on the hebephrenic cluster. The excited and the hebephrenic patients are similar with respect to another scale, number 29 — they both tend to violate rules and regulations. It is not surprising that there should be some similarity between the excited and the hebephrenic type of schizophrenic clusters because the scales contain certain items in common. As a matter of fact, in a newly admitted sample of patients in whom the deteriorated or hebephrenic schizophrenic symptoms are not yet manifest, it is probable that the excited and the hebephrenic cluster scores would show the same pattern of relationship. The hebephrenic schizophrenic cluster score is also correlated with some of

[3]This interpretation is confirmed by a scrutiny of individual ratings.

the scores which are significantly correlated with the depressed cluster score. They are: number 23, do not tend to feel that people are against them, and number 25, do not have an unrealistic delusional or grandiose notion concerning their role or status.

## SUMMARY AND CONCLUSIONS

The occupational and practical significance of the empirically determined symptom clusters are studied by correlating each of nine symptom cluster scores with each of 61 occupational rating scales. The sample comprises a group of 60 schizophrenic patients at the Greystone Hospital, Greystone Park, New Jersey. With the limitations which this sample imposes the following conclusions are offered:

1. The occupational ratings for patients who have high scores on the manic cluster, on the excited cluster, on the anxiety cluster, on the depressed cluster, and on the hebephrenic cluster are all conspicuously and meaningfully different from each other in a manner congruent with the general interpretation given these scores. The two cluster scores which resemble each other most with respect to their occupational scale correlations are the excited and hebephrenic schizophrenia scores.

2. Patients having relatively high scores on the conversion hysteria cluster and on the phobic-compulsive cluster are not characterized by the ratings they receive on the occupational rating scales. No general significance may be ascribed to this finding, however, because the present sample comprises patients who for the most part have almost no conversion symptoms and relatively few phobic-compulsive symptoms.

3. In the present sample, many patients who are quite paranoid with respect to both the paranoid schizophrenia cluster score and the paranoid condition score are not characterized by any of the occupational disadvantages expressed by the occupational rating scales. It is suggested, therefore, that the condition of the paranoid patient and the severity and social significance of his illness should not be evaluated on the basis of his performance in occupational therapy or other special services of mental hospitals. His danger to himself and to society lies not in his occupational limitations, but lies in the nature of his conception of his role relative to that of his fellow men. His false conceptions can be readily discerned by an examining psychiatrist and may be rated.

*Received March 30, 1951.*

## REFERENCES

1. WITTENBORN, J. R. A new procedure for evaluating mental hospital patients. *J. consult. Psychol.*, 1950, **14**, 500–501.
2. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, **15**, 290–302.
3. WITTENBORN, J. R., AND HOLZBERG, J. D. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, **15**, 372–380.
4. WITTENBORN, J. R., AND HOLZBERG, J. D. The Wechsler-Bellevue and descriptive diagnosis. *J. consult. Psychol.*, 1951, **15**, 325–329.
5. WITTENBORN, J. R. AND HOLZBERG, J. D. The Rorschach and descriptive diagnosis. *J. consult. Psychol.*, 1951, **15**,
6. WITTENBORN, J. R., AND LESSER, G. S. Biographical factors and psychiatric symptoms. *J. clin. Psychol.*, 1951, **7**, 317–322.
7. WITTENBORN, J. R., HOLZBERG, J. D., AND SIMON, B. Symptom correlates for descriptive diagnosis. *Genetic Psych. Monograph.*

# A NOTE ON FOSBERG'S "FOUR EXPERIMENTS WITH THE SZONDI TEST"

JACOB COHEN

NEW YORK UNIVERSITY SCHOOL OF EDUCATION AND BRONX V. A. HOSPITAL

AN excellent series of experiments with the Szondi [2] was unfortunately marred by an error in reasoning involving the chance expectancy of varying numbers of selections in a diagnostic category of the Szondi. The error does not invalidate the conclusions of the study, but since an increasing amount of research is being done with this instrument, it is perhaps worth while to set the record straight and thus prevent repetition of the error.

Fosberg evaluates the chance expectancy of zero through six "like" and "dislike" choices in a given Szondi factor by means of expanding the appropriate binomials. He reasons, quite correctly, that since two "liked" photographs are chosen in each set, the probability of any factor being thus chosen is 2 in 8 or ¼. He states, however, that the probability of a factor being chosen as "disliked" is 2 in 6, or 1/3, apparently reasoning that following the selection of two photographs as "liked," only six remain from which to choose the two "disliked." This leads to the conclusion that it is more probable that a given photograph be chosen as "disliked" (1/3) than as "liked" (¼). This seems wrong intuitively.

If we reason somewhat differently [1], namely, that irrespective of "like" or "dislike," four photographs are chosen from eight, and half of the four are "liked" and half "disliked," we arrive at the same probability for both (which seems more reasonable), namely $p = ¼$.

The error in [2] is this: It is true that when two photographs are selected as "disliked," only six remain (two having been previously selected as "liked"), thus yielding the value 2/6 or 1/3. However, for a photograph to be present among the remaining six, it must not have been chosen as "liked." The probability of the latter is 3/4 (1 minus ¼, the probability of selection as "liked"). Thus, the simultaneous occurrence of the photograph *not* being selected as "liked," and *being* selected as "disliked" is 3/4 times 1/3 which equals ¼, the same probability as was obtained for "liked" (as in [1]).

Fosberg's error led to incorrect values for the chance expectancies of zero through six "dislike" choices in a factor and further resulted in mistaken conclusions in chi square tests of the divergence of two empirically obtained distributions from these chance expectancies [2, Table 3]. The correct theoretical chance distribution for "dislikes" is the same as that for "likes," correctly stated in [2, p. 39]. When the chi square values are computed using the correct distribution, the following changes in results are found:

1. From significantly different from chance to not significant: *s*— and *e*— in the normal group, and *p*— and *d*— in the psychiatric group.

2. From not significantly different from chance to significant: *hy*— and *k*— in the psychiatric group.

The present author would like to stress the fact that these corrections in no way vitiate Fosberg's conclusions and knows of no objective experimental evidence which can challenge his statement that, "It may well be that there are psychological forces at work in determining the selection of pictures, but they are probably not what Szondi workers believe them to be" [2, p. 40].

*Received April 11, 1951.*

REFERENCES

1. COHEN, J. The theoretical distribution of Szondi valences. *J. consult. Psychol.*, 1951, **15**, 130–133.
2. FOSBERG, I. A. Four experiments with the Szondi test. *J. consult. Psychol.*, 1951, **15**, 39–44.

# NEW BOOKS AND TESTS

## Books

BENNETT, GEORGE K., SEASHORE, HAROLD G., AND WESMAN, ALEXANDER G. *Counseling from profiles, a casebook for the Differential Aptitude Tests.* New York: Psychological Corp., 1951. Pp. 95. $1.75.

Counselors in nine school systems provided thirty typical case studies of eighth-grade to twelfth-grade pupils who had been counseled with the aid of the DAT. The attractive booklet presents these case studies, each with comments by the authors of the tests, as an aid to counselor training. While the case studies are not assumed to prove the validity of the tests, they certainly demonstrate the value of a differentiated profile, as opposed to a single scholastic aptitude score.

BERNARD, HAROLD W. *Toward better personal adjustment.* New York: McGraw-Hill, 1951. Pp. vii + 439. $4.00.

An orientation textbook, intended to guide college students toward better mental health and personal efficiency. Its psychological theory is antiquated. The topic of learning, for example, is treated almost entirely in terms of William James's precepts about habits, and "good" and "bad" habits are dichotomized freely. As a practical tool for college guidance, however, the book is not without assets. The author admirably avoids the temptation to mystify his readers or to frighten them with morbid cases. Optimism and realism are blended appropriately, and the emphasis on self-responsibility for personal improvement is constructive.

CAMERON, NORMAN, AND MAGARET, ANN. *Behavior pathology.* Boston: Houghton Mifflin, 1951. Pp. xvi + 645. $5.00.

Let no one underestimate this book — it is one of the most original and significant contemporary contributions to the understanding of human behavior. With skill and creativity, the authors have integrated the central recent insights about deviant behavior: that it originates from learning, that the learning takes place in social contexts and involves communication and role-taking, and that psychotherapy is social relearning. The book inevitably invites comparison with the senior author's *The Psychology of Behavior Disorders* (1947). The present volume is not a revision or replacement of the earlier one. Instead, the two books are supplementary. *Behavior Disorders,* as its name implies, was organized around the major syndromes. *Behavior Pathology* is organized in terms of the person and his behavior; from beginning to end it is truly a psychology with little trace of descriptive psychiatry remaining. For example, there are chapters on pseudocommunity, on autistic community, on disorganization, and on desocialization as behavior phenomena, but no chapter on schizophrenia as such. The two books differ in a much more important respect, however. *Behavior Disorders* tended to be one-sided. It presented the author's biosocial theory in relative isolation from other contemporary thinking. *Behavior Pathology* is a work of greater scope, that incorporates relevant evidence on needs, learning processes, conflict, and anxiety, and yet retains its essential unity. All psychologists will find this book stimulating to their thinking, whether their interests lie in clinical applications or in more general psychological theory. The volume will also be a useful textbook. It does not fit the conventional outline of the course in abnormal psychology, but instructors daring enough to reorganize their course around it will probably find that the adventure is rewarding to their students as well as to themselves.

CARROLL, HERBERT A. *Mental hygiene, the dynamics of adjustment.* (2nd Ed.) New York: Prentice-Hall, 1951. Pp. xii + 448. $5.65.

The revision of Carroll's *Mental Hygiene* is somewhat expanded from the first edition, and about one-third of its material is new. The chapters on motivation and learning have been revised extensively, with varying success. Except for a good discussion of anxiety, the treatment of motivation is largely enumerative. The discussion of learning is greatly improved and cites recent and pertinent experimental studies. The book remains a highly desirable text for undergraduate instruction. Its shortcomings with respect to organization and precision are offset by the merits of its simple writing, its objective viewpoint, and its genuinely hygienic attitudes.

DAVIDOFF, EUGENE, AND, NOETZEL, ELINOR S. *The child guidance approach to juvenile delinquency.* New York: Child Care Publications, 1951. Pp. vi + 173. $4.50.

This is a loosely written, poorly organized collection of papers distantly related to the title of the

NOTE: Some reviews in this issue were prepared by the Associate Editors, who may be identified by their initials. Unsigned reviews are by the Editor. —*L. F. S.*

book. Beginning with a most obvious discussion of the "organization of the child guidance clinic" and the functions of psychiatrists, psychologists, and social workers, the authors proceed to discuss the treatment of different "types" of delinquents. Much of this discussion consists of long lists of "types," of "do's" and "don't's," and of illustrative cases in which the "dynamics" are intricately and unconvincingly spun out. Several disconnected chapters follow on delinquents and drama techniques; on the treatment of pathological liars (with a list of 36 types of liars); on "additional therapeutic considerations"; and on a "follow-up study of 374 children," among other chapters. The follow-up consisted essentially of determining whether these children were or were not subsequently registered with the Social Service Exchange, and little else. The case for the child guidance clinic in the treatment of juvenile delinquents is not enhanced by this volume.—*M. K.*

DEWEY, RICHARD, AND HUMBER, W. J. *The development of human behavior.* New York: Macmillan, 1951. Pp. xv + 762. $5.50.

A sociologist and a psychologist collaborated in preparing this interestingly integrated text, suitable for introducing undergraduates to the social sciences. The book's five sections deal with foundations of psychological and social theory, development from infancy to old age, deviant roles and statuses, social institutions, and the problem of the good society. The degree of integration is a real achievement. Although mature scholars can readily label most of the subject matter, students will perceive the book as "one course," and will forever after be fortified against spurious compartmentalizations. Also, the style is lively and readable, and the illustrative material well chosen. It is unfortunate that the book is marred by some evidences of carelessness — a mislabeled figure, a garbled reference — but these remediable shortcomings do not entirely conceal its general excellence.

GILMER, B. VON HALLER. *How to help your child develop successfully.* New York: Prentice-Hall, 1951. Pp. viii + 368. $3.95.

A book for parents, which gives questions and answers on typical parent-child problems, sketches development from birth to ten years, and provides a 122-page section for recording a child's physical and psychological development. The author takes a middle course between "training" and indulgence, and writes with an open-minded good humor that will help many parents and alarm few. Advocates of the current trend toward full indulgence will not like the book; other psychologists who are no more in accord with the excesses of Ribble than with those of Watson will recommend it. The style of writing is breezy and appropriately nontechnical, and the numerous illustrations are amusing.

HARROWER, M. R., AND STEINER, M. E. *Large scale Rorschach techniques.* (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1951. Pp. xx + 353. $8.50.

The second edition of the manual for the Group Rorschach and the Multiple Choice Test is not a revision in the usual sense. The main body of the original book is unchanged, save for the deletion of numerous pages of superfluous graphs, and for a few minor editorial corrections. The principal new material is confined to a 53-page added section on "recent developments," which does scant justice to its topic. Of the 34 "new" studies cited, 26 are dated 1947 or earlier. Moreover, the section's poor organization gives evidence of hasty and careless preparation. A really up-to-date annotated bibliography would have been more serviceable. There is also a supposedly new section on "card-pull" that turns out to be a rearrangement of some old data on the relative frequencies of locations, determinants, and content categories. The second edition does not give nearly as good support as it might to the genuinely useful Group Rorschach. On the other hand, it maintains the unduly optimistic attitude of the first edition toward the Multiple Choice Test, in spite of the bulk of adverse evidence.

HARTWELL, SAMUEL W. *Citizens' handbook of sexual abnormalities and the mental hygiene approach to their prevention.* Washington, D. C.: Public Affairs Press, 1951. Pp. viii + 71. $1.00.

One of the reports of the Michigan Commission on the Deviated Criminal Sex Offender is this helpful booklet for laymen, which takes up the types of sex deviations, common fallacies, and the preventive mental hygiene of sex. The prevention of sex deviations is discussed mainly in terms of the psychoanalytic interpretation of parent-child relationships.

JAHODA, MARIE, DEUTSCH, MORTON, AND COOK, STUART W. *Research methods in social relations.* Vol. 1: *Basic processes.* Vol. 2: *Selected techniques.* New York: Dryden Press, 1951. Pp. x + 422, x + 337. $3.75 each vol., $6.00 set.

This able reference work and textbook on social research methods was developed by a committee of the Society for the Psychological Study of Social Issues. The first volume, written mainly by Jahoda, Deutsch, and Cook, takes up the major steps of scientific inquiry: the formulation of a problem, research design, measurement, data collection, analysis and interpretation, and presentation and application. Volume two consists of eleven chapters by fifteen principal contributors that deal with some of the more specific technical problems of social research, such as constructing questionnaires, interviewing, observing, sample design, and scaling. Although the discussion of designs and techniques is focused mainly on the problem of prejudice, the material has broader applicability. Because the research problems and tools of social psychology are so close

to those of clinical psychology, the present work is probably the best available approximation to a handbook for clinical research. It will become a widely used resource for graduate students struggling with dissertations, for hard-pressed professors, and for clinical research workers generally.

JONES, ARTHUR J. *Principles of guidance and pupil personnel work.* (4th Ed.) New York: McGraw-Hill, 1951. Pp. xvi + 630. $4.75.

The author states that this edition "is not essentially different from the third in general organization and in fundamental concepts." He views guidance as an aspect of personnel work, but argues for the continued use of the concept. Proportionally great space is devoted to the need, purpose, administration, and method of investigation in guidance. Special chapters describe guidance in elementary schools, in junior high schools, in secondary schools, in colleges, with out-of-school youth, and with Negro youth. Although Jones mentions recent literature in the field, his whole frame of reference seems to the reviewer to reflect the emphasis at the time of his first edition rather than that of today. One might expect that, in 1951, more space would be devoted to the dynamics of human behavior which is to be guided rather than to the importance and machinery of guidance as such.—*F. McK.*

MARX, MELVIN H. (Ed.) *Psychological theory: contemporary readings.* New York: Macmillan, 1951. Pp. xi + 585. $5.00.

If the maturity of a science can be judged by the systematic and critical attention that it pays to the construction of its theories, then psychology has indeed matured greatly in the past two decades. In what is perhaps the first book of its kind, Marx has effectively documented the contemporary growth of psychological theory. The well-selected papers are divided into two groups: the first twenty-three deal directly with problems of psychological theory construction; the remaining twenty-four illustrate the process from the writings of important recent psychological theorists. The total impression made by the papers is very different from that conveyed by the typical survey of "schools of psychology" that was so popular twenty or more years ago. There are no warring antagonisms. Instead, the contemporary viewpoints when placed side by side seem highly complementary to one another. One does not ask whether molar or molecular concepts, or behavioristic or field-theoretical psychologies, are "right" or "wrong," but sees instead their distinctive contributions toward a common understanding. The collection will be an indispensable reference for advanced students.

MICHIGAN, STATE OF. *Report of the Governor's Study Commission on the Deviated Criminal Sex Offender.* Detroit, Mich.: Governor's Study Commission, 1951. Pp. xii + 245.

A two-year study by a commission appointed by the governor of Michigan led to a number of broad and constructive recommendations of procedures for dealing with sex deviates. Notable proposals are for continued and extended research, for the provision of differentiated treatment facilities, for indeterminate sentences for treatable offenders, for greater consideration of the mental health of the victims of sex crimes, for the education of police and other public officers, and for preventive measures through parent-child education. The recommendations are supported by statistical data, by case studies, and by reports of the deliberations of subcommittees.

MUDD, EMILY HARTSHORNE. *The practice of marriage counseling.* New York: Association Press, 1951. Pp. xx + 336. $4.50.

A survey of the development and status of marriage counseling, with special reference to the work of the Philadelphia Marriage Council. The practical chapters on the functions and processes of counseling, with numerous case studies, provide solid training materials for persons preparing to be marriage counselors and for general counselors who encounter marital problems. There is an especially good discussion of the relationships of the various professional groups that deal with aspects of marriage. An appendix gives condensed reports of the resources and practices of thirty-five marriage counseling agencies.

POWERS, EDWIN, AND WITMER, HELEN. *An experiment in the prevention of delinquency.* New York: Columbia Univ. Press, 1951. Pp. xliii + 649. $6.00.

This volume is a report of the Cambridge-Somerville Youth Study, an action research designed to test the hypothesis that juvenile delinquency can be prevented by establishing a sustained, friendly relationship between a boy and a counselor who can give him a satisfactory ego ideal. Although the original experimental design was a carefully contrived one, changing personnel and the effects of the war years disrupted the standards that had been set up. However, it was felt wise to report the results, even if incomplete. The study used two groups of boys. Each experimental boy had all the benefits of the counselor and of the study's resources, while his control "twin" was influenced only by the ordinary events and situations that he encountered in his culture. In the report, Powers describes the design and the findings and Witmer, in the role of a critical outside observer, evaluates the successes and failures of the study. Gordon W. Allport contributes a challenging foreword. Not least among the authors are the boys themselves, who give their own reactions to the experiment. This refreshingly frank and objective report is a "must" for all who are concerned with juvenile delinquency.—*B. M. L.*

SCHWEBEL, MILTON. *The interests of pharmacists.* New York: King's Crown Press, 1951. Pp. xii + 84. $1.75.

In addition to developing a pharmacist scale for the Strong Vocational Interest Blank, this study explored the relationship of job satisfaction to interest measurement. Its techniques and findings are applicable to the measurement of interest and satisfaction in other occupations.

SIGEL, DAVID. *Frustration in adolescent youth.* Federal Security Agency, Office of Education, Bulletin 1951, No. 1. Washington: U. S. Government Printing Office, 1951. Pp. vi + 65. 25¢.

This pamphlet about adolescents and their problems is addressed primarily to educators and school counselors. Its theoretical orientation draws about equally from the phenomenological approach of Snygg and Combs and from Maier's analysis of reaction to frustration as behavior without a goal. Except for a probably unjustified faith in questionnaires as instruments for identifying frustrated persons, the practical recommendations for prevention and rehabilitation seem to be well conceived.

STRECKER, EDWARD A., EBAUGH, FRANKLIN G., AND EWALT, JACK R. *Practical clinical psychiatry.* (7th Ed.) Philadelphia: Blakiston, 1951. Pp. xii + 506. $7.00.

This is the seventh, revised edition of what has now become a leading textbook in the psychiatric field. The authors' names will attest to the soundness of the treatment, and the revision is a genuine and extensive one. It can be recommended to students and to practicing clinicians who may want available a reference book with a more medical orientation than is found in the usual psychological treatment of the mental disorders.—*W. A. H.*

## TESTS

ALLPORT, GORDON W., VERNON, PHILIP E., AND LINDZEY, GARDNER. *Study of Values, Revised Edition.* Test booklets ($2.50 per 25) with manual, pp. 16. Boston: Houghton Mifflin, 1951.

After twenty years of wide and varied use, the *Study of Values* has been revised thoroughly. There are changes both in content and in format. The old "social" value, which had proved unreliable because of adherence to Spranger's excessively broad definition, has been redefined in terms of altruistic or philanthropic interest. Most of the items have been revised, to simplify wordings, and to eliminate outdated and overspecific cultural references. The new items give the impression that they are likely to maintain their meanings with the passage of time. The manual gives an extensive report of the statistical characteristics of the revision. New and revised items were selected by three successive item analyses, and each scale is internally consistent. Scale reliabilities are substantially improved, ranging from .87 to .92 on a one-month retest. Scale intercorrelations are mainly low. Norms are supplied in terms of the means and *SD*'s of 1,816 college students, and are further analyzed in terms of sex and of occupational groups. The new format simplifies the process of scoring and eliminates the need for a separate score sheet. In its excellent new revision, the *Study of Values* seems destined to several more decades of service as a tool for guidance and for research.

MOORE, BRUCE V., LAPP, C. J., AND GRIFFIN, CHARLES H. *Engineering and Physical Science Aptitude Test, Revised Manual.* New York: Psychological Corp., 1951. Pp. 11.

The *Engineering and Physical Science Aptitude Test,* a 72-minute composite of parts on mathematics, formulation, physical science comprehension, arithmetic reasoning, verbal comprehension, and mechanical comprehension, was developed during World War II as a placement battery for accelerated engineering programs. The revised manual contains validation data, intercorrelations, and percentile norms for college and employee groups.

ROEDER, WESLEY S., AND GRAHAM, HERBERT B. *Aptitude Tests for Occupations.* Grade 9-adult. 6 tests, 1 form of each: Personal-Social, Mechanical, General Sales, Clerical Routine, Computational, Scientific. 12 (20 )to 20 (23) min. each test, 107 (125) min. total. Test booklets (10¢ ea., $1.25 to $2.00 per 25, according to test), with manual, key; profile sheet (5¢); specimen set, 1 test (35¢). Los Angeles: California Test Bureau, 1951.

The *Aptitude Tests for Occupations* series consists of six test booklets, one for each of the areas covered. The Personal-Social and General Sales tests use described situations with multiple-choice questions. The Mechanical test employs pulley systems, pattern layouts, and tool recognition. Clerical Routine includes name and number checking and spelling. The Computational test uses arithmetical calculations and estimates, and the Scientific test consists mainly of reasoning and comprehension with scientific materials. Full data are given on reliabilities, which run mainly from .78 to .91. The test intercorrelations are moderate, mostly from .40 to .70 for males, and lower for females. The correlations of the tests with other similar tests are stated in considerable detail and are, on the whole, satisfactory. Percentile norms for each test are tabulated for each grade from 9 to 13, and separately for males and females. In the main, these tests fulfill the current expectations of psychologists for a series to measure differentiated abilities. They fall decidedly short of ultimate ideals for test development in two respects. First, there is no evidence that they measure "aptitude" as distinguished from achievement, and the magical word is likely to deceive the unwary. Second, although the tests

are recommended mainly for occupational guidance, they are not accompanied by evidence to show their power to predict successful adjustment to any vocation. The latter fault can be remedied by further research.

SULLIVAN, ELIZABETH T., CLARK, WILLIS W., AND TIEGS, ERNEST W. *California Short-Form Test of Mental Maturity, 1950 S-Form.* 5 levels: grades Kgn-1, 1-3, 4-8, 7-10, 9-adult. 1 form at each level. 20 (30) to 52 (70) min., according to level. Test booklets ($1.75 per 25) with manual, key, record sheet; Scoreze answer sheet (7¢); IBM answer sheet (4¢); specimen set, 1 level (35¢). Los Angeles: California Test Bureau, 1950.

The revision of this group mental ability test places more stress on the differential aptitudes measured, and less on the total MA and IQ. The test for each stage of maturity consists of seven parts, presumably comparable from level to level because of similar content. Scores are obtained for total mental factors, language factors, nonlanguage factors, spatial relationships, logical reasoning, numerical reasoning, and verbal concepts. Total score reliabilities range from .92 to .95, and part-score reliabilities from .81 to .95, being generally higher at the upper age levels. The subtest intercorrelations, mainly from the .20's to the .40's, give some support to the differential interpretation of the profiles. Norms are based on very large groups that were controlled with respect to age and school progress. No evidence is given, however, about the geographic or socioeconomic distribution of the normative subjects. Validity is defended in terms of high correlation with the Stanford-Binet, but the exact coefficient is not stated. No data are presented on the test's relationships to subsequent school achievement or to other appropriate criteria. In these respects, the information given in the manuals falls short of meeting desirable professional standards.

WELLS, F. L. *Modified Alpha Examination, Form 9, Revised Manual.* New York: Psychological Corp., 1951. Pp. 8.

The revised manual for this ancient but still occasionally useful warhorse includes revised norms for high school pupils and for 1,331 adult male engineering employees.

WESMAN, ALEXANDER G. *Wesman Personnel Classification Test, Revised Manual.* New York: Psychological Corp., 1951. Pp. 15.

The *PCT* is a test for business and industrial use that yields separate verbal and numerical ability scores. The 1951 edition of the manual reports a number of studies of the test's applications, and gives additional norms.

# INDEX TO VOLUME 15

## SUBJECT INDEX


ACE, and personality inventory data, 160
Adjustment and values, index of, 257
Adlerian therapy, 32
Anxiety, problem of, in client-centered therapy, 359
Attitudes, toward self and others, 79

Behavior, psychologist's, as area of research, 278
Book reviews, 85, 167, 263, 354, 445, 512

Case reporting, psychological, at psychiatric staff conferences, 425
Cerebral palsied children, handedness in, 9
Children, normal, situational play therapy with, 225
Client-centered therapy, 72
  problem of anxiety in, 359
Client responses, and counseling method, 216
Clinical interns, procedures and methods in supervising, 267
Clinical psychologist, and mental defectives, 1
Clinical psychology, oscillation and progress in, 281
  trainees, supervision of in neuropsychiatric hospital, 271
  where are we going, 171
Color, effects of, on thematic fantasy, 412
Conferences, psychiatric staff, psychological case reporting at, 425
Counseling method, and client responses, 216
Counselor, general, eclecticism of, 76

Deterioration, in senile psychosis, 145
Drinkers, psychometric study of, 501
Drive theory, and TAT responses, 45

Einstellung-effect, as test of rigidity, 89
Einstellung test, of rigidity, relation to concreteness of thinking, 303
Electroconvulsive therapy, effect of, on psychological test performances, 430
Experimental stress, Rorschach factors related to, 190

Faking, experimentally induced in TAT, 469
Frustration, induced, 18
  induced in thematic apperception stories, 400

Group conformity rating, Rosenzweig Picture-Frustration Study, 250
Guidance center, community, supervision of interns in, 268

Handedness, in cerebral palsied children, 9
Harrower-Erickson multiple choice test, patients with spinal cord injuries, performance on, 346
Hewson ratios, 163
Humm-Wadsworth, and Minnesota Multiphasic, 457

Insight, two inventories of, 109
Intelligence, epileptic, 392
Intellectual classification, semantics of, 487
Intelligence level, Rorschach "organic signs" and, 343
Interns, in community guidance center, supervision of, 268
Interview devices, projective pictures as, 405
I.P.A.T. Culture-Free Intelligence Scale 2
  IQ standardization of, 154
IQ, variability, in relation to age, 231
  on revised Stanford-Binet, 231

Language, schizophrenic, regression of, 387

Masculinity-femininity measures, comparison of, 484
Mental defectives, and clinical psychologist, 1
Mental hospital, patients, symptoms patterns in group of, 290
Military personnel, nervous symptoms among, 320
Minnesota Multiphasic Personality Inventory, and Humm-Wadsworth, 457
  factor analysis of fifteen scales of, 451
  internal structure of, 134
  item analysis aid for, 262
  norms, reliability and forms, 435
  profile coding, 82
  profiles in multiple sclerosis, 253
  profiles in patients with low-back pain, 350
  role-playing ability test derived from, 102
  *Sc* scale, 142
Multiple sclerosis, MMPI profiles in, 253

Nervous symptoms, among military personnel, relation to age, combat experience, and marital status, 320
Neuropsychiatric hospital, clinical psychology trainees in, supervision of, 271
Nondirective therapy, 32
Normals, young, and schizophrenics, young, performance on Wisconsin Card Sorting Test, 311

Personality, self-evaluative questionnaires as projective measures of, 496
Projective measures of personality, self-evaluative questionnaires, 496
Projective pictures, as interview devices, 405
Projective technique, Michigan Picture Test, 51
  see also Rorschach Szondi; TAT
Psychiatric symptoms, practical correlates of, 505
Psychiatric syndromes, generality of, 372
Psychoanalysis, transference in, 72
Psychoanalytic therapy, 32
Psychotherapy, accuracy of self-evaluations in, 95
  group, evaluation of by means of performance tests, 363
  hearing versus reading interviews, 115
  and insight, 109
  and role-playing ability, 102
  training in, 274

Psychotics, organic, symbol arrangement test, 439
Punishment-reward expectancy in thematic apperception stories, 400

Research, psychologist's behavior as an area of, 278
Research scientists, psychological tests of, 492
Rigidity, Einstellung-effect as test of, 89
Einstellung test of, relation to concreteness of thinking, 303
Role-playing ability, test of, 102
Rorschach and Behn-Rorschach, inkblot tests, comparison of, 185
cards, color as a determinant of responsiveness to, in schizophrenia, 196
color and 8-9-10 per cent, 122
descriptive diagnosis, 460
factors related to experimental stress, 190
forerunner of, 120
impulsivity as indicated by, 464
organic signs and intelligence level, 343
norms, a contribution to, 178
norms for "shock," 127
predicting continuation in therapy, 368
test results, influence of set for area on, 337
Rosenzweig Picture-Frustration study, conformity rating of, internal consistency of, 250

Schizophrenic language, regression of, 387
Schizophrenics, young, and normals, young, performance on Wisconsin Card Sorting Test, 311
Scientists, research, psychological tests of, 492
Self-evaluations, accuracy of, 95
Self-evaluation questionnaires as projective measures of personality, 496
Senile psychoses, relation of writing speed to, 243
Situational play therapy, with normal children, 225
Stanford-Binet, and Wechsler Intelligence Scale for Children, 236, 330
and Wechsler Intelligence Scale for Children, pupil functioning on, 475
IQ variability on, 231
Symbol arrangement, original test of, validated on organic psychotics, 439
Symptom patterns, in group of mental hospital patients, 290
Syphilitics, congenital, performance of, on Wechsler Intelligence Scale for Children, 239
Szondi test, instability score, development of, 334
four experiments with, 39
note on Fosberg's four experiments with, 511
pictures, factors affecting the choice of, 210
pictures, stimulus values of, 419
valence distribution, 130

Test, ACE, 160
Einstellung-effect and rigidity, 89
Einstellung, of rigidity, relation to concreteness of thinking, 303
Harrower-Erickson multiple choice, performance of patients with spinal cord injuries, 346
Humm-Wadsworth, 457
I.P.A.T. Culture-Free Intelligence Scale, 2, 154
Michigan Picture, 51
MMPI, 82, 102, 134, 142, 350, 435, 457
performance and evaluation of group psychotherapy, 363
performance, psychological, effect of brief stimulus electroconvulsive therapy on, 430
Rorschach, 120, 122, 127, 178, 185, 190, 196, 337, 343, 368, 460, 464
Szondi, 39, 130, 210, 334, 419, 505
Stanford-Binet, 231, 236, 330, 469
symbol arrangement, 439
Thematic Apperception, 45, 55, 203, 469
Wechsler-Bellevue 68, 151, 163, 325
Wechsler Intelligence Scale for Children, 62, 236, 239, 330, 381, 385, 475
Wisconsin Card Sorting, 311
Thematic Apperception Test, and drive theory, 45
card, ambiguity values of, 203
experimentally induced faking of responses, 469
two methods of administration, 55
Thematic fantasy, some effects of color on, 412
Thematic apperception stories, induced frustration and punishment-reward expectancy in, 400
Therapist, reaction of patient to, used as training and research device, 24
Therapy, Adlerian
brief stimulus electroconvulsive, effect of on psychological test perfomance, 430
client-centered, 72
client-centered, problem of anxiety in, 359
nondirective, 32
prediction continuation in, by Rorschach test, 368
psychoanalytic, 32, 72
reaction of patient to therapist in, 24
situational play, with normal children, 225
Transference, in client-centered and psychoanalytic therapy, 72

Voice, and clinical judgments, 115

Wechsler-Bellevue, and descriptive diagnosis, 325
neurotics and schizophrenics, 151
schizophrenia, 68
validity of Hewson ratios, 163
Wechsler Intelligence Scale for Children and Stanford-Binet, 236, 330, 475
equivalent test and mental ages for, 381
evidence on validity of, 385
performance of congenital syphilitics on, 239
Verbal and Performance IQ's, 62
Wisconsin Card Sorting Test, 311
Writing Speed, relation to age and senile psychoses, 243

# AUTHOR INDEX

The names of original articles are printed in CAPITALS; and of authors of books and tests reviewed in Roman.


ABORN, M., 231
ABRAMSON, L. S., 337
Abt, L. E., 85
ALESSI, S. L., 425
ALLEN, R. M., 262
Allport, G. W., 167, 515
ANDERSON, G. L., 267
ANDREW, G., 51
Ann Arbor Conference on Training of Psychological Counselors, 357
ARBUCKLE, D. S., 76
Arelrad, S. W., 446

Babcock, H., 263
Bauman, M. K., 354
Beach, F. A., 355
BEIER, E. G., 359
Bellak, L., 85
BENARICK, S. J., 142
Bennett, G. K., 448, 449, 512
BERGMAN, D. V., 216
Bergman, M., 263
Bernard, H. W., 512
BEST, H. L., 419
BIJOU, S. W., 203
BILLS, R. E., 257
Birch, H. G., 87
BOREN, J. E., 145, 243
Blake, R. R., 445
Bleuler, E., 167
BOTWINICK, J., 145, 243
BRACKBILL, G. A., 412
Brayfield, A. H., 85
BREWER, J. E., 268
Brosse, T., 167
BROWN, C. C., 343
Brun, R., 445
BUKER, S. L., 196
Burling, T., 87
Burnham, P. S., 354

Cameron, N., 512
CANTER, A. H., 253
Cantril, H., 167
CARP, A. L., 271
Carroll, H. A., 512
Cass, M. T., 354
Cass, W. A., JR., 178
CATTELL, R. B., 154
Cattell, R. B., 167
CLARK, G. C., 262
Clark, W. W., 516
Cohen, J., 130, 511
Colby, K. M., 445
COLGIN, R., 435
COLLINS, A. L., 392
Cook, S. W., 513
CORSINI, R., 487
COTZIN, M., 1
Council of State Governments, 170
CRANDALL, V. J., 400
Crawford, L. E., 449
Crawford, P. L., 167
Crow, A., 445
Crow, L. D., 445

DAVID, H. P., 334
Davidoff, E., 512
Davis, O. C., 446
DE HAAN, H., 68
Dennis, W., 168
DERNER, G. F., 231
Deutsch, M., 513
Devereux, G., 263
Dewey, R., 513
DIEPPA, J., 469
DIERS, W. C., 343
DiMichael, S. G., 263
DOLL, E. E., 9
Dollard, J., 168
Dorcus, R. M., 85
Dumpson, J. R., 167
Dunsmoor, C. C., 446

Ebaugh, F. G., 515
EICHLER, R. M., 185
Eissler, R. S., 168
ELIAS, G., 496
Ellis, A., 354
ELLSWORTH, R. B., 387
ERON, L. D., 45, 55
Ewalt, J. R., 515

Ferenczi, S., 85
FEY, E. T., 311
FIEDLER, F. E., 32
Finegold, S. N., 446
Fischer, S., 446
Fisher, V. E., 86
Fodor, N., 86
Ford, C. S., 355
FOSBERG, I. A., 39
FRANDSEN, A. N., 236
Freeman, L., 446
Fromm-Reichmann, F., 86

GARDNER, R. W., 464
Garrison, K. C., 263
Gaynor, F., 86
GEORGE, C. E., 190
Gilbert, G. M., 86
GILLILAND, A. R., 435, 457
Gilmer, B. v. H., 513
Ginsberg, S. W., 446
Ginzberg, E., 446
Goheen, H. W., 355
Gordon, L. V., 170
Graham, H. B., 515
Greco, M. C., 86
Griffin, C. H., 515
GROSSMAN, D., 109
Guetzkow, H., 446
Guilford, J. P., 86
Gulliksen, H., 168
GURVITZ, M. S., 120
GUTHRIE, G. M., 142, 385

HAMMOND, K. R., 368
HAMPTON, P. J., 501
HANVIK, L. J., 350
Harris, R. E., 86
Harrower, M. R., 513
Hartley, E. L., 87
Hartley, R. E., 87
HARTWELL, S. W., 51
Hartwell, S. W., 513
Hathaway, S. R., 355, 449
Hayes, S. P., 354
Hellersberg, E. F., 168
HEPPELL, H. K., 405
Herma, H., 169
Herma, J. L., 446
HIGGINSON, J. B., 236
HOLT, R. R., 95
HOLZBERG, J. D., 325, 372, 425, 460
Horney, K., 169
Horrocks, J. E., 355
Humber, W. J., 513
Hunt, J. McV., 87
HUTT, M. L., 51

Jahoda, M., 513
Jones, A. J., 514
JONES, F. D., 363
Josiah Macy, Jr. Foundation, The, 266
JUSTMAN, J., 475

KAHN, T. C., 439
Kasius, C., 355
Kavruck, S., 355
KENNY, D. T., 203
Kitson, H. D., 446
KNAUSS, J., 368
Kogan, L. S., 87
Kohn, N., Jr., 447
KRUGMAN, J. I., 475
KRUGMAN, M., 475
Kuder, G. F., 449
KUHLEN, R. G., 320
Kurth, G. M., 169

LAMPHIEAR, D. E., 419
Lapp, C. J., 515
Lefever, D. W., 87
LEHNER, G. F. J., 134
Leiter, R. G., 88
Lemert, E. M., 355
Liberman, S., 356
Lindquist, E. F., 263
LINDZEY, G., 18
Lindzey, G., 515
LITTLE, K. B., 134
Loewenstein, R. M., 264
Lorand, S., 169, 264
Lovell, C., 86
Lovell, H. W., 447
LUCHINS, A. S., 24, 89, 303
LUFT, J., 115

Maas, H. S., 447
McCLELLAND, W. A., 102
McHugh, G., 88
McKinley, J. C., 449
McLEAN, O. S., 257
McREYNOLDS, P., 178
Magaret, A., 512
Malamud, D. I., 167
Marx, M. H., 514
Maslow, A. H., 356
Meehl, P. E., 355
MERRILL, M. A., 281
METTLER, F. A., 505
Michigan, State of, Goverior's Study Commission, 514
Millbank Memorial Fund, 266
Miller, J. G., 86
Miller, N. E., 168
Mittelmann, B., 356
Mons, W., 447
Mooney, R. L., 170
Moore, B. V., 515
MOSAK, H. H., 346
MOUSTAKAS, C. E., 225
MOWRER, O. H., 274
Mowrer, O. H., 264
Mudd, E. H., 514
Muench, G. A., 86
Munn, N. L., 447
Myklebust, H. R., 169

Naumburg, M., 87
Noetzel, E. S., 512
NOLLER, P. A., 330

Palmer, S. H., 354
Parry, J. B., 88
Partington, J A., 88
PASTOVIC, J. J., 385
Pear, T. H., 356
PEMBERTON, C. L., 160
PERLMAN, J. A., 122
PETERS, H. N., 363
PHILLIPS, E. L., 79
PITTS, V. A., 239
Porter, E. H., Jr., 264
Powers, E., 514

RABINOWITZ, W., 334
RAIMY, V. C., 405
Raimy, V. C., 264
Ramsey, G. V., 356, 445
Rapaport, D., 169, 447
Reitman, F., 447
Rennie, T. A. C., 87
RIECKEN, H. W., 18
RITTER, A. M., 55
ROE, A., 492
Roeder, W. S., 515
ROGERS, C. R., 171
Rogers, C. R., 265
ROGERS, L. S., 151, 368
Roheim, G., 87

ANDERSON, H., 127
ARASON, S. B., 278
CHERER, I. W., 430
Schilder, P., 170, 265
Schneiders, A. A., 447
SCHRAMM, T. A., 330
Schwebel, M., 515
SEASHORE, H. G., 62
Seashore, H. G., 449, 512
Sechehaye, M. A., 448
Segel, D., 515
Selover, M., 87
Shaffer, G. W., 85
SHEPLER, B. F., 484
Shock, N. W., 356
Shore, M. J., 170
Sifferd, C. S., 170
Smith, G. E., 448
SMITH, S., 190
SNYDER, W. U., 142
Steiner, M. E., 513
Stevens, S. S., 448
Stolzenberg, J., 265
Stone, L. J., 86
Strang, R., 448
Strecker, E. A., 515
Strunk, M., 167
Sullivan, E. T., 516
SULLOSI, E., 419

TAYLOR, M. V., JR., 250
TAYLOR, O. M., 250
Teuber, H. L., 86
Thorndike, R. L., 356
Thorne, F. C., 87
Thurstone, L. L., 170
Tiegs, E. W., 516
Turrell, A. M., 87
TYLER, F. T., 451

VANCE, E. L., 257
Vernon, P. E., 88, 357, 515

WALKER, M. S., 9
WALLEN, R., 210
WALTON, R. E., 51
WEBB, W. B., 68
WECHSLER, D., 381
WEIDER, A., 330
WEISSKOPF, E. A., 469
Weitzel, H. I., 87
Wells, F. L., 516
WELSH, G. S., 82
Wepman, J. M., 265
Wesman, A. G., 449, 512, 516
WEXLER, M., 425
WHEELER, J. I., JR., 163
WHEELER, W. M., 134
WILKINS, W. L., 163
WILLIAMS, M., 196
Witmer, H., 514
WITTENBORN, J. R., 45, 290, 325, 372, 460, 505
Witty, P., 357
Wolff, W., 88
WOOD, A. B., 72
Woodward, L. E., 87
WRIGHTSTONE, J. W. 475

Yacorzynski, G. K., 266
YOUNG., F. M., 239

Zangwill, O. L. 448
Zubin, J., 86

## PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

### Volume 64, 1950

**Patterns of Personality Rigidity and Some of Their Determinants.** Seymour Fisher Elgin State Hospital. #307, $1.00

**The Value of an Oral Reading Test for Diagnosis of the Reading Difficulties of College Freshmen of Low Academic Performance.** Charles A. Wells, American International College. #308, $1.00

**Rorschach Responses Related to Vocational Interests and Job Satisfaction.** Solis L. Kates, Michigan State College. #309, $1.00

**Symbol Elaboration Test (S.E.T.): The Reliability and Validity of a New Projective Technique.** Johanna Krout, Chicago Psychological Institute. #310, $2.00

**Changes in Responses to the Minnesota Multiphasic Inventory Following Certain Therapies.** William Schofield, University of Minnesota. #311, $1.00

**A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport.** Carroll H. Leeds, Furman University. #312, $1.00

**The Nature and Efficacy of Methods of Attack on Reasoning Problems.** Benjamin Burack, Roosevelt College. #313, $1.00

**The Validity of a Multiple-Choice Projective Test in Psychopathological Screening.** Martin Singer, VA Hospital, Long Island. #314, $1.00

**A Normative Study of the Thematic Apperception Test.** Leonard D. Eron, Yale University. #315, $1.50

**Experimentally Induced Variations in Rorschach Performance.** Edith E. Lord, Arizona State Department of Health. #316, $1.00

**An Evaluation of Personality-Trait Ratings Obtained by Unstructured Assessment Interviews.** Ernest C. Tupes, United States Air Force. #317, $1.00

The 1950 volume of the Psychological Monographs consists of eleven separate issues. Orders for any of these can be placed separately at the prices listed above, or the entire volume can be ordered for $6.00.

## American Psychological Association

**1515 Massachusetts Avenue N.W., Washington 5, D.C.**

# JOURNAL OF CONSULTING PSYCHOLOGY
# AVAILABLE BACK ISSUES

The issues listed below are available for sale

| YEAR | VOLUME | JAN FEB | MAR APR | MAY JUN | JUL AUG | SEP OCT | NOV DEC | PRICE PER NUMBER | PRICE PER VOLUME |
|---|---|---|---|---|---|---|---|---|---|
| 1937 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1938 | 2 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1939 | 3 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1940 | 4 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1941 | 5 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1942 | 6 | 1 | – | 3 | – | – | 6 | $.60 | $1.80 |
| 1943 | 7 | – | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1944 | 8 | 1 | – | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1945 | 9 | 1 | 2 | 3 | 4 | 5 | 6 | $.60 | $3.00 |
| 1946 | 10 | 1 | 2 | 3 | 4 | 5 | – | $.60 | $3.00 |
| 1947 | 11 | – | – | 3 | 4 | 5 | 6 | $.60 | $2.40 |
| 1948 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | $1.00 | $5.00 |
| 1949 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | $1.00 | $5.00 |
| 1950 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | $1.00 | $5.00 |
| 1951 | 15 | By Subscription, $5.00 | | | | | | | |

**Price for Volumes 1 through 14 $46.20**

**Table based on inventory of January 1, 1951**

The Journal of Consulting Psychology is of particular interest to clinical psychologists, psychiatrists, school psychologists, and persons who are engaged in counseling and guidance work.

The price per volume of the Journal of Consulting Psychology was $3.00 between 1937 and 1947. It is now $5.00 per volume and $1.00 per issue. For foreign orders, $.25 per volume should be added. Address orders to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 Massachusetts Ave. N. W.
Washington 5, D. C.

# NEUROLOGY & PSYCHIATRY

**Section VIII of Excerpta Medica**

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neuro-physiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition, the following special branches are dealt with fully:

* **Psychoanalysis, individual and analytical psychology**
* **Sexology, criminology, alcoholism, and drug addiction as related to psychiatry**
* **Social and industrial psychology and psychiatry, vocational guidance and mental testing**
* **Psychometrics**
* **Personality testing and the Rorschach test**
* **Heredity and statistical studies as they apply to psychiatry**
* **Mental defect and epilepsy**

Annual volume of 1000 pages, including an index classified by author and subject.

**Subscription price $22.50 per annual volume**

**THE WILLIAMS & WILKINS COMPANY**

Agents for the United States, Canada and Central America

**Mt. Royal and Guilford Aves.** **Baltimore 2, Maryland**

---

Please enter my subscription to Section VIII, NEUROLOGY AND PSYCHIATRY, OF EXCERPTA MEDICA. I enclose $22.50.

Name.....

Address.....

.....

New, original ideas highlighted the now-famous Hixon Symposium on

# CEREBRAL MECHANISMS IN BEHAVIOR

Here's the published record of that Symposium edited by LLOYD A. JEFFRESS
*The University of Texas*

Some of the men at the Hixon Symposium contributed papers, some augmented the discussion by brief remarks, some were on a discussion panel — all are widely known leaders in their fields. In contributing their stimulating ideas, the members struck at the periphery of a number of fields — psychologists will find this record of the Symposium of immense value and interest. The book consists of six major papers, each followed by a discussion section, two additional discussion sections, and a final evaluation.

**Send for an approval copy**

**The Papers**

The General and Logical Theory of Automata, *by* DR. JOHN VON NEUMANN. Why the Mind Is in the Head, *by* DR. WARREN S. MCCULLOCH. The Problem of Serial Order in Behavior, *by* DR. K. S. LASHLEY. Functional Differences between the Occipital and Temporal Lobes, *by* DR. HEINRICH KLUVER. Relational Determination in Perception, *by* DR. WOLFGANG KOHLER. Brain and Intelligence, *by* DR. WARD HALSTEAD. The Symposium from the Standpoint of a Clinician, *by* DR. HENRY W. BROSIN.

**December 1951.**
**311 pages**
**$6.50**

**JOHN WILEY & SONS, Inc., 440 Fourth Ave., New York 16, N. Y.**

## PHILOSOPHICAL LIBRARY PUBLICATIONS

**SANTAYANA: Atoms of Thought**
A judicious anthology of pithy and worthwhile thoughts of this, our foremost idealist philosopher. $5.00

**EINSTEIN: Out of My Later Years**
Here is Einstein the philosopher, scientist, and man . . . an independent and uncompromising thinker, unencumbered by traditional prejudices, inhibitions and limitations in the world of today. $4.75

**DICTIONARY OF PHILOSOPHY**
*by Dagobert D. Runes*
"A handy volume, all-embracing in content, clear in exposition, objective in viewpoint and earmarked by a correctness that is inescapable."—Dean James F. Carroll, Duquesne University. $6.00

**A HISTORY OF PHILOSOPHICAL SYSTEMS**
*Edited by Vergilius Ferm*
The first comprehensive survey of all leading schools and movements in philosophy, treated by 41 eminent scholars. $6.00

**BERTRAND RUSSELL**
*by H. W. Leggett*
Russell's philosophy, his approach to mathematics, his views on marriage, world affairs, religion, morality and sex. Illustrated $3.75

**PSYCHOLOGY AND ITS BEARING ON EDUCATION**
*by C. W. Valentine, Univ. of Birmingham*
The topics include: The Development and Training of Personality and Character; The Basic Motives; Suggestion; Unconscious Influences; Sex Education; Learning and Remembering; Repression and Discipline; Play and Activity Methods; The Interests of Children; The Acquisition of Skill; Training in Reasoning; General Intelligence and Special Abilities, and their testing; Estimating Personality and Character; Educational and Vocational Guidance; School Records; Stages of Development in Infancy, Middle Childhood and Adolescence; Backward, Problem and Delinquent Children. A brief appendix gives simple explanation of the most essential statistical methods applied to psychology and education. $6.00

**Expedite Shipment by prepayment**

**PHILOSOPHICAL LIBRARY, PUBLISHERS**
**15 East 40th Street Desk 195 New York 16, N.Y.**

# THE NORTHAMPTON ACTIVITY RATING SCALE

**Isidor W. Scherer, Ph.D.**
CHIEF CLINICAL PSYCHOLOGIST
V. A. HOSPITAL, NORTHAMPTON, MASSACHUSETTS

---

The Northampton Activity Rating Scale is a new clinical method of recording the kinds of behavior exhibited by mental patients in rehabilitation and activity therapies of the hospital environment.

The ratings obtained through the use of these scales are appropriate for many kinds of evaluation, such as:

1. **The effectiveness of various rehabilitation activities for various kinds of patients.**
2. **The practical consequences of behavioral changes following other therapies such as electroconvulsive shock or psychosurgery.**
3. **The relationship between psychological test findings and actual behavior in real-life situations.**
4. **Vocational rehabilitation possibilities of the patient.**
5. **Standardization of criteria employed when a patient is considered for discharge.**
6. **Evidence concerning the manner in which the patient relates to different features of a situation including the personalities of occupational therapists or other supervisors.**
7. **Criteria used in the study of many types of prognostication that are made for the patient.**

RELIABILITY: The scales are reliable to such a degree that in 1242 pairs of independent ratings none of the pairs of ratings differed from each other by more than two points. The reliability coefficients range from .62 to .99+ with a median value of .94.

**Specimen set including manual, 30 cents.**

---

**Published and distributed by**
**MEED SCIENTIFIC APPARATUS COMPANY**
**Box 658, Springfield, Massachusetts**